

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2002, 00:12:53 ; Search time 1474.82 Seconds
(without alignments)
5190.255 Million cell updates/sec

Title: US-09-621-781-1_COPY_1507_1970

Perfect score: 464

Sequence: 1 gaaagagaagaagtggg.....tttgcaaccgacggac 464

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result		Query			Description	
No.	Score	Match	Length	DB	ID	
1	464	100.0	1940	10	AF351130	AF351130 Rattus no
2	40	8.6	2088	10	CRUGADD34X	L28147 Hamster Gad
3	35	7.5	2275	10	NMMDPRMR	X51829 Mouse myelo
4	35	7.5	215734	2	AC073828	AC073828 Mus muscu
5	23	5.0	216	10	NMUB3984	U83984 Mus musculu
6	23	5.0	224	10	CGU83983	U83983 Citicetus
7	22	4.7	185414	2	AL355526	AL355526 Homo sapi
8	22	4.7	190856	2	AC024505	AC024505 Homo sapi
9	21	4.5	61147	9	AL162294	AL162294 Human DNA
10	21	4.5	155190	2	AC026323	AC026323 Homo sapi
11	21	4.5	165228	9	AC005325	AC005325 Homo sapi
12	20	4.3	1776	8	MTLPATP9S	Z50200 L.perenne m
13	20	4.3	2805	4	AF286474	AF286474 Bos tauru
14	20	4.3	108893	9	HSDJ130E4	AL078582 Human DNA
15	20	4.3	110000	2	PFMAL4P1_0	AL034557 Plasmodiu
16	20	4.3	159644	9	AP000820	AP000820 Homo sapi
17	20	4.3	163186	9	AB045360	AB045360 Homo sapi
18	20	4.3	165079	2	AC024517	AC024517 Homo sapi
19	20	4.3	167300	2	AC021394	AC021394 Homo sapi
20	20	4.3	171326	2	AL353894	AL353894 Homo sapi
21	20	4.3	175504	2	AL591916	AL591916 Homo sapi
22	20	4.3	180615	2	AC025549	AC025549 Homo sapi
23	20	4.3	197926	2	AC073844	AC073844 Homo sapi
24	20	4.3	300000	9	AP002533	AP002533 Homo sapi
25	19	4.1	371	11	HSC009YH5	Z67713 H.sapiens D
26	19	4.1	8372	14	AF017780	AF017780 Sour cher
27	19	4.1	72928	6	AR103481	AR103481 Sequence
28	19	4.1	82918	8	AB010071	AB010071 Arabidops
29	19	4.1	97142	2	AC010439	AC010439 Homo sapi
30	19	4.1	114022	9	AL157952	AL157952 Human DNA
31	19	4.1	115391	2	AC022439	AC022439 Homo sapi
32	19	4.1	126936	2	AC090472	AC090472 Homo sapi
33	19	4.1	126956	2	AC090730	AC090730 Homo sapi
34	19	4.1	151037	9	AC018741	AC018741 Homo sapi
35	19	4.1	151673	10	AC087102	AC087102 Rattus no
36	19	4.1	151970	2	AC040905	AC040905 Homo sapi
37	19	4.1	162030	2	AC018601	AC018601 Homo sapi
38	19	4.1	167859	2	AC016391	AC016391 Homo sapi
39	19	4.1	171850	2	AC021405	AC021405 Homo sapi
40	19	4.1	178671	2	AC069283	AC069283 Homo sapi
41	19	4.1	186448	2	AC034129	AC034129 Homo sapi
42	19	4.1	190174	2	AC022739	AC022739 Homo sapi
43	19	4.1	191756	2	AL512785	AL512785 Homo sapi
44	19	4.1	202851	2	AC015684	AC015684 Homo sapi
45	19	4.1	206454	2	AC011227	AC011227 Homo sapi

ALIGNMENTS

RESULT 1
AF351130
LOCUS AF351130 1940 bp DNA
DEFINITION Rattus norvegicus progression elevated gene-3 protein (Peg-3) gene,
Promoter region.
ACCESSION AF351130
VERSION AF351130.1 GI:13562027
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 1940)
AUTHORS Su,Z., Shi,Y. and Fisher,P.B.
TITLE Cooperation between API and PEA3 sites within the progression
elevated gene-3 (PEG-3) promoter regulate basal and differential
expression of PEG-3 during progression of the oncogenic phenotype
in transformed rat embryo cells
JOURNAL Oncogene 19 (30), 3411-3421 (2000)

MEDLINE 20374129
PUBMED 10918598
REFERENCE 2 (bases 1 to 1940)
AUTHORS Su, Z., Shi, Y. and Fisher, P.B.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2001) Urology, Columbia University, 630 West 168th Street, New York, NY 10032, USA

FEATURES
source
1. 1940
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
promoter
1. 1940
/gene="Peg-3"
1. >1940
/gene="Peg-3"
/note="progression elevated gene-3"
BASE COUNT 568 a 496 c 468 g 408 t
ORIGIN

Query Match 100.0%; Score 464; DB 10; Length 1940;
Best Local Similarity 100.0%; Pred. No. 5.9e-239;
Matches 464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaagagaagaagaatggagacagcatgtgactgctgatatgaattggcgtgtgtcctca 60
Db 1477 GAAAGAGAAGAAGATGGAGACAGCATGTGACTGCCCTGATGAAGTTGGCGTGTCTCAA 1536

QY 61 aagttctgcagatgtgacgctctctgattgagccaagagacacgctctgggaagccacg 120
Db 1537 AAGTTCTGCGAGATTGACGGCTCTCTGGATTGTGAGCCCAAGACACGCCCTGGGAAGCCACG 1596

QY 121 gtgacctcacaagagcccggaatctccgaggaatttcagttgttttctctctccacc 180
Db 1597 GTGACCTCACAAAGGCCCGGAATCTCCGGAGAAATTTCAGTGTGTTTCTCTCTCCACC 1656

QY 181 ttctcagggacttcgcaaacctccgctctccggtgacgtcagcatagcgtctgcacga 240
Db 1657 TTTCTCAGGGACTTCCGAACCTCCGCCCTCTCCGGTGACGTACGATAGCGCTGCTCAGA 1716

QY 241 ctataaactcccggtgctgctgtgtggcgagattgactcagttcgcagctgttggaga 300
Db 1717 CTATAAATCCCGGGTGTGTTGGCGCAGATTGACTCAGTTCCGACGCTTGTGGAAGA 1776

QY 301 ttacatgcgagaccccgcgagactccgcatccctttccgggagacagcctttgcagacc 360
Db 1777 TTACATGCGAGACCCCGCGACTCCGCATCCCTTTGCCGGGACAGCCCTTTGCCAGACC 1836

Y 361 cgtgagacatcacgtcccccagagcccaagcgtgagggcgacatgaacgctgtgacctga 420
Db 1837 CGTGAGACATCACGTCCCGGAGCCCGCCTTGAGGGCGACATGAACGCGCTGCGCTTGA 1896

QY 421 gagcaatccggaccacgactcgttttgccaacccgaaccggac 464
Db 1897 GAGCAATCCGGACCCACGATCGCTTTGGCAAACCGAACC GGAC 1940

RESULT 2
CRUGADD34X 2088 bp mRNA ROD 29-JUN-1994
LOCUS Hamster Gadd34 mRNA, complete cds.
DEFINITION
ACCESSION L28147
VERSION L28147.1 GI:452489
KEYWORDS acidic protein.
SOURCE Cricetulus longicaudatus cDNA to mRNA.
ORGANISM Cricetulus longicaudatus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.

REFERENCE 1 (bases 1 to 2088)
AUTHORS Zhan, Q., Lord, K.A., Alamo, I.Jr., Hollander, M.C., Carrier, F., Ron, D., Kohn, K.W., Hoffman, B., Liebermann, D.A. and Fornace, A.J.Jr.
TITLE The gadd and myd genes define a novel set of mammalian genes

JOURNAL encoding acidic proteins that synergistically suppress cell growth
MOL. CELL. BIOL. 14, 2361-2371 (1994)
MEDLINE 94187707
FEATURES
source
location/Qualifiers
1. 2088
/organism="Cricetulus longicaudatus"
/db_xref="taxon:10030"
/cell_line="Chinese hamster ovary-K1"
215. 1987
/gene="Gadd34"
215. 1987
/gene="Gadd34"
/note="Gadd34 is the hamster homolog of the mouse MyD16 cDNA; contains multiple PEST sequences implicated in rapid protein turnover"
/codon_start=1
/protein_id="AAA36983.1"
/db_xref="GI:452490"
/translation="MAPSPRPOHLLWRDAHSFHLSPILMGFLSRWSRLRVPEAPWP
PAETVTGADQIEADAHAPAPPLVPENHPQGEAEESGTPEEGKAAOGPCLDVQANSS
PETLGLSDDDKOGDGPREGRAHAGLPILSPGLQADKSLGEEVAGEGVTETLAY
PTSHWEGCPSSEEDDEGETVKKAFRASADSPGHKSSTSVYCPGEAEHQATEKOTENK
DPPSSPSGSHSRAMEYCSKOEGADPEPRAGKYOLCONAEAESEAEKVSLSVSSG
NAFLKAWVYRPGEDTDDDDSDWGSAAEEGKALSSPTSPENHFLKAWVYRPGEDT
DDSDWGSAAEEGKALSSPTSPENHFLKAWVYRPGEDTDDSDWGSAAEEGKALSSPT
TPHTSAFLKTYVCCPGEDTDDDCVYVPEDESAADPKSBSHAQGLPEQTEGLV
EAEHSLFQVAEYFLPGEKPAWPWTAPKPLRLRLRQLTLRLTPQDODPEPLRARKVH
SENVTHFLAVWAGPAQAARGPWEQLARDRSFARRIAQAEKILGPLYLPAPFARAW
ARLGNPSLPALPEPICDHTFFPSQ"

repeat_region 1035. 1154
repeat_region 1155. 1274
repeat_region 1275. 1372
repeat_region 1406. 1457
polya_signal 2066. 2071
BASE COUNT 487 a 602 c 594 g 405 t
ORIGIN

Query Match 8.6%; Score 40; DB 10; Length 2088;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 tcacgtcccgagccccacgctgagggcgacatgaagc 409
Db 79 TCACGTCCCGAGCCCGCAGCGCTGAGGGCGACATGAACGC 118

RESULT 3
MMMDPRMR
LOCUS MMMDPRMR 2275 bp mRNA ROD 12-SEP-1993
DEFINITION Mouse myeloid differentiation primary response mRNA encoding MyD16 protein.
ACCESSION X51829
VERSION X51829.1 GI:53040
KEYWORDS MYD16 gene; myeloid differentiation primary response gene.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 2275)
AUTHORS Liebermann, D.A.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-1990) Liebermann D.A., University of Pennsylvania School of Medicine, Dept of Biochemistry & Biophysics, Philadelphia PA 19104-6059, USA
REFERENCE 2 (bases 1 to 2275)
AUTHORS Lord, K.A., Hoffman, B. and Liebermann, D.A.
TITLE Sequence of MyD16 cDNA: a novel myeloid differentiation primary response gene induced by IL6
JOURNAL Nucleic Acids Res. 18 (9), 2823 (1990)

MEDLINE 90251472 Location/Qualifiers
FEATURES
source
1. 2275
/organism="Mus musculus"
/strain="SL"
/db_xref="taxon:10090"
/clone_lib="lambda ZAP"
/clone="116"
229. .2202
/note="MyD16 protein (AA 1-657)"
/codon_start=1
/protein_id="CAA36128.1"
/db_xref="GI:53041"
/db_xref="SWISS-PROT:P17564"
/translation="MAPSPRPQHLHWDAHNFYLLSPLMGLSRAMRRLGPEVPEA
WLAKVTGADQIEAALLTPVSGNLLPHGETEESGSPESQAQRCLVEAESPP
ETWGLSNDEYNAPGQDDLREKEMERTAGKATLQAGLQAGADKRLGEVVAAREGVAE
PAYPTSQLEGPAENEDEGETVKTQASAA5IAPGKPTVPFPLGEAEHQATEEKG
ENKADPSNSPSSGSHSRAWEYYSREKPKQEGEAKVEAHKSTGHPCRMNAEEGPE
TFVCTGNAFILKAWYRPGEDTEEDNSDSAEEDTAQTGATPHTSAFLKAWYRPG
DTEEDSDSDSAEDTAQTGATPHTSAFLKAWYRPGEDTEEDNSDSAEEDTAQT
ATPHTSAFLKAWYRPGEDTEEDNSDSAEEDTAQTGATPHTSAFLKAWYRPGEDT
EDDTEEDSESNVAPGDETDSSQSPCLQPRCLPGDKTGRGEEPLFOYAFYLP
EKDESPWAPKPLPLRLRRLFLKAPTRDQDPEIPLKARKVHFAEKTIVHFLAVWAGP
AQARRGPWEQFARDRSRARRIAQAEKILGPYLTPLDSRARAWARLRNP5LPQSEPRS
SEATPLTQDVTTPSP1PSETPSP5LYLGRRG"
2275
polya_site
/note="polya site"
BASE COUNT 577 a 652 c 636 g 410 t
ORIGIN
Query Match 7.5%; Score 35; DB 10; Length 2275;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 373 cgtcccgagcccaagcctgagggcgacatgaac 407
|||||
Db 96 CGTCCCGAGAGCCCGAGCGCTGAGGGGACATGAAC 130
RESULT 4
AC073828 215734 bp DNA HTG 29-JUN-2000
LOCUS
DEFINITION Mus musculus clone RP23-9J18, WORKING DRAFT SEQUENCE, 21 unordered
pieces.
AC073828 AC073828.1 GI:8810445
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1740562
Center clone name: RPCI-23_9J18

Summary Statistics
Consensus quality: 196667 bases at least Q40

Consensus quality: 207971 bases at least Q30
Consensus quality: 209928 bases at least Q20
Estimated insert size: 207740; agarose-fp estimation
Estimated insert size: 213734; sum-of-contigs estimation
Quality coverage: 6.12 in Q20 bases; agarose-fp estimation
Quality coverage: 5.94 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1013: contig of 1013 bp in length
* 1014 1113: gap of unknown length
* 1114 2175: contig of 1062 bp in length
* 2176 2275: gap of unknown length
* 2276 3429: contig of 1154 bp in length
* 3430 3529: gap of unknown length
* 3530 4884: contig of 1355 bp in length
* 4885 4984: gap of unknown length
* 4985 6848: contig of 1864 bp in length
* 6849 6949 9317: gap of unknown length
* 9318 9417: contig of 2369 bp in length
* 9418 12128: gap of unknown length
* 12129 12228: contig of 2711 bp in length
* 12129 12228: gap of unknown length
* 12229 14970: contig of 2742 bp in length
* 14971 15070: gap of unknown length
* 15071 19571: gap of 4501 bp in length
* 19572 19671: gap of unknown length
* 19672 25961: contig of 6290 bp in length
* 25962 26061: gap of unknown length
* 26062 36753: gap of 10692 bp in length
* 36754 36853: gap of unknown length
* 36854 49116: contig of 12263 bp in length
* 49117 49216: gap of unknown length
* 49217 61879: contig of 12663 bp in length
* 61880 61979: gap of unknown length
* 61980 72513: gap of 10534 bp in length
* 72514 72613: gap of unknown length
* 72614 83721: contig of 11108 bp in length
* 83722 83821: gap of unknown length
* 83822 98633: contig of 14812 bp in length
* 98634 98733: gap of unknown length
* 98734 118323: contig of 19590 bp in length
* 118324 118423: gap of unknown length
* 118424 137668: contig of 19245 bp in length
* 137669 137768: gap of unknown length
* 137769 162338: contig of 24570 bp in length
* 162339 162438: gap of unknown length
* 162439 186625: contig of 24187 bp in length
* 186626 186725: gap of unknown length
* 186726 215734: contig of 29009 bp in length.

FEATURES
source
1. 215734
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-9J18"
/clone_lib="RPCI mouse BAC library 23"

BASE COUNT 54325 a 53055 c 52767 g 53583 t 2004 others
ORIGIN

Query Match 7.5%; Score 35; DB 2; Length 215734;
Best Local Similarity 100.0%; Pred. No. 6.8e-08;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 373 cgtcccgagcccaagcctgagggcgacatgaac 407
|||||
Db 8415 CGTCCCGAGAGCCCGAGCGCTGAGGGGACATGAAC 8449

RESULT 5
LOCUS MMU83984 216 bp DNA ROD 07-JUL-1998
DEFINITION Mus musculus apoptosis associated protein (GADD34) gene, promoter sequence.
ACCESSION U83984
VERSION U83984.1 GI:3258621
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 216)
AUTHORS Hollander, M.C., Zhan, Q., Bae, I. and Fornace, A.J. Jr.
TITLE Mammalian GADD34, an apoptosis- and DNA damage-inducible gene
JOURNAL J. Biol. Chem. 272 (21), 13731-13737 (1997)
MEDLINE 97298078
REFERENCE 2 (bases 1 to 216)
AUTHORS Hollander, M.C. and Fornace, A.J. Jr.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1997) LMP, NIH, Bldg. 37, Room 5C09, Bethesda, MD 20892, USA

FEATURES
source location/Qualifiers
1..216
/organism="Mus musculus"
/db_xref="taxon:10090"
<1..>216
/gene="GADD34"
/note="growth arrest and DNA damage-inducible gene 34; Myd16 gene"
<1..>216
promoter /gene="GADD34"
135..146
/gene="GADD34"
149..158
/gene="GADD34"
/note="ATF/CRE site"
177..182
TATA_signal /gene="GADD34"
40 a 66 c 63 g 47 t
BASE COUNT
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 23; DB 10; Length 216;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 203 ccgcctccggtgacgtcagca 225
|||||
Db 138 CCGCCTCTCCGGTGACGTACGCA 160

RESULT 6
LOCUS CGU83983 224 bp DNA ROD 07-JUL-1998
DEFINITION Cricetulus griseus apoptosis associated protein (GADD34) gene, promoter sequence.
ACCESSION U83983
VERSION U83983.1 GI:3258620
KEYWORDS
SOURCE Chinese hamster.
ORGANISM Cricetulus griseus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus. 1 (bases 1 to 224)
AUTHORS Hollander, M.C., Zhan, Q., Bae, I. and Fornace, A.J. Jr.
TITLE Mammalian GADD34, an apoptosis- and DNA damage-inducible gene
JOURNAL J. Biol. Chem. 272 (21), 13731-13737 (1997)
MEDLINE 97298078
REFERENCE 2 (bases 1 to 224)
AUTHORS Hollander, M.C. and Fornace, A.J. Jr.

TITLE Direct Submission
JOURNAL Submitted (07-JAN-1997) LMP, NIH, Bldg. 37, Room 5C09, Bethesda, MD 20892, USA

FEATURES
source location/Qualifiers
1..224
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
<1..>224
/gene="GADD34"
/note="growth arrest and DNA damage-inducible gene 34"
<1..>224
promoter /gene="GADD34"
135..146
/gene="GADD34"
149..158
/gene="GADD34"
/note="ATF/CRE site"
174..179
TATA_signal /gene="GADD34"
46 a 67 c 63 g 48 t
BASE COUNT
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 23; DB 10; Length 224;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 203 ccgcctccggtgacgtcagca 225
|||||
Db 138 CCGCCTCTCCGGTGACGTACGCA 160

RESULT 7
LOCUS AL355526/c 185414 bp DNA HTG 20-JAN-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-152L7, *** SEQUENCING IN PROGRESS ***
ACCESSION AL355526
VERSION AL355526.4 GI:9797321
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 185414)
AUTHORS Plumb, B.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 12, 2000 this sequence version replaced gi:9231038.

COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba152L7
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 173494 bases at least Q40
Consensus quality: 177549 bases at least Q30
Consensus quality: 180091 bases at least Q20
Insert size: 183314; sum-of-contigs
Insert size: 188970; 9.8% error; agarose-fp
Quality coverage: 4.22x in Q20 bases; sum-of-contigs Quality coverage: 4.28x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 3619: contig of 3619 bp in length
* 3620 3719: gap of 100 bp
* 3720 8736: contig of 5017 bp in length
* 8737 8836: gap of 100 bp
* 8837 13069: contig of 4233 bp in length
* 13070 13169: gap of 100 bp
* 13170 42396: contig of 29227 bp in length
* 42397 42496: gap of 100 bp
* 42497 113471: contig of 70975 bp in length
* 113472 113571: gap of 100 bp
* 113572 119459: contig of 5888 bp in length
* 119460 119559: gap of 100 bp
* 119560 122149: contig of 2590 bp in length
* 122150 122249: gap of 100 bp
* 122250 125172: contig of 2923 bp in length
* 125173 125272: gap of 100 bp
* 125273 129826: contig of 4554 bp in length
* 129827 129926: gap of 100 bp
* 133356 133355: contig of 3429 bp in length
* 133356 133455: gap of 100 bp
* 133456 136798: contig of 3343 bp in length
* 136799 136898: gap of 100 bp
* 136899 140265: contig of 3367 bp in length
* 140266 140365: gap of 100 bp
* 140366 142977: contig of 2612 bp in length
* 142978 143077: gap of 100 bp
* 143078 146062: contig of 2985 bp in length
* 146063 146162: gap of 100 bp
* 146163 149355: contig of 3193 bp in length
* 149356 149455: gap of 100 bp
* 149456 154688: contig of 5233 bp in length
* 154689 154788: gap of 100 bp
* 154789 162417: contig of 7629 bp in length
* 162418 162517: gap of 100 bp
* 162518 170626: contig of 8109 bp in length
* 170627 170726: gap of 100 bp
* 170727 175237: contig of 4511 bp in length
* 175238 175337: gap of 100 bp
* 175338 177559: contig of 2222 bp in length
* 177560 177659: gap of 100 bp
* 177660 180444: contig of 2785 bp in length
* 180445 180544: gap of 100 bp
* 180545 185414: contig of 4870 bp in length.

FEATURES

source

1. .185414
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-152L7"
/clone_lib="RPC1-11.1"
1. .3619
/note="assembly_fragment:01810
fragment_chain:1"
3720. .8736
/note="assembly_fragment:01729
fragment_chain:1"
8837. .13069
/note="assembly_fragment:00298
fragment_chain:1"
13170. .42396
/note="assembly_fragment:00928
fragment_chain:1
clone_end:SP6
vector_side:left"
42497. .113471
/note="assembly_fragment:01958
fragment_chain:1"
113572. .119459

/note="assembly_fragment:00442
fragment_chain:2"
119560. .122149
/note="assembly_fragment:00656
fragment_chain:2"

misc_feature
/note="assembly_fragment:01717
fragment_chain:2"

misc_feature
/note="assembly_fragment:00504
fragment_chain:3"

misc_feature
/note="assembly_fragment:02152
fragment_chain:3"

misc_feature
/note="assembly_fragment:00586
fragment_chain:4"

misc_feature
/note="assembly_fragment:00823
fragment_chain:4"

misc_feature
/note="assembly_fragment:00123"
143078. .146062

misc_feature
/note="assembly_fragment:00837"
146163. .149355

misc_feature
/note="assembly_fragment:00879"
149456. .154688

misc_feature
/note="assembly_fragment:01030"
154789. .162417

misc_feature
/note="assembly_fragment:01278"
162518. .170626

misc_feature
/note="assembly_fragment:01660"
170727. .175237

misc_feature
/note="assembly_fragment:01758"
175338. .177559

misc_feature
/note="assembly_fragment:01777"
177660. .180444

misc_feature
/note="assembly_fragment:01980"
180545. .185414

misc_feature
/note="assembly_fragment:02113"
BASE COUNT 57924 a 34224 c 34805 g 56330 t 2131 others

Query Match 4.7%; Score 22; DB 2; Length 185414;
Best Local Similarity 100.0%; Pred. No. 0.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gaaagagaagaatggagaca 22
Db 175922 GAAAGAGAAAGAAATGGAGACA 175901

RESULT 8
AC024505/c LOCUS AC024505 190856 bp DNA HTG 12-MAY-2000
DEFINITION Homo sapiens chromosome 1 clone RP11-152L7 map 1, WORKING DRAFT
SEQUENCE, 29 unordered pieces.

ACCESSION AC024505
VERSION AC024505.3 GI:7770482
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 190856)
AUTHORS Birren,B., Linton,L., Nussbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 1, clone RP11-152L7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 190856)
AUTHORS Birren,B., Linton,L., Nussbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,

TITLE
JOURNAL
COMMENT

Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Laroque,K., Lehocsky,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGuirk,A., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivar,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Titrrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wymann,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.

Direct Submission

Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 12, 2000 this sequence version replaced gi:7547164.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L7197

Center clone name: 152_L_7

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 172979 bases at least Q40

Consensus quality: 182209 bases at least Q30

Consensus quality: 185744 bases at least Q20

Insert size: 190000; agarose-fp

Insert size: 188056; sum-of-contigs

Quality coverage: 3.7 in Q20 bases; agarose-fp

Quality coverage: 3.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1005: contig of 1005 bp in length
* 1006 1105: gap of 100 bp
* 1106 2321: contig of 1216 bp in length
* 2322 2421: gap of 100 bp
* 2422 4167: contig of 1746 bp in length
* 4168 4267: gap of 100 bp
* 4268 6402: contig of 2135 bp in length
* 6403 6502: gap of 100 bp
* 6503 9001: contig of 2499 bp in length
* 9002 9101: gap of 100 bp
* 9102 11665: contig of 2564 bp in length
* 11666 11765: gap of 100 bp
* 11766 13958: contig of 2193 bp in length
* 13959 14058: gap of 100 bp
* 14059 17065: contig of 3007 bp in length
* 17066 17165: gap of 100 bp
* 17166 20430: contig of 3265 bp in length
* 20431 20530: gap of 100 bp
* 20531 24620: contig of 4090 bp in length
* 24621 24720: gap of 100 bp

FEATURES
source

* 24721 28004: contig of 3284 bp in length
* 28005 28104: gap of 100 bp
* 28105 31520: contig of 3416 bp in length
* 31521 31620: gap of 100 bp
* 31621 35687: contig of 4067 bp in length
* 35688 35787: gap of 100 bp
* 35788 40442: contig of 4655 bp in length
* 40443 40542: gap of 100 bp
* 40543 48006: contig of 7464 bp in length
* 48007 48106: gap of 100 bp
* 48107 55271: contig of 7165 bp in length
* 55272 55371: gap of 100 bp
* 55372 61182: contig of 5811 bp in length
* 61183 61282: gap of 100 bp
* 61283 68644: contig of 7362 bp in length
* 68645 68744: gap of 100 bp
* 68745 75906: contig of 7162 bp in length
* 75907 76006: gap of 100 bp
* 76007 83998: contig of 7992 bp in length
* 83999 84098: gap of 100 bp
* 84099 93489: contig of 9391 bp in length
* 93490 93589: gap of 100 bp
* 93590 101953: contig of 8364 bp in length
* 101954 102053: gap of 100 bp
* 102054 110821: contig of 8768 bp in length
* 110822 110921: gap of 100 bp
* 110922 121798: contig of 10877 bp in length
* 121799 121898: gap of 100 bp
* 121899 133647: contig of 11749 bp in length
* 133648 133747: gap of 100 bp
* 133748 146812: contig of 13065 bp in length
* 146813 146912: gap of 100 bp
* 146913 160242: contig of 13330 bp in length
* 160243 160342: gap of 100 bp
* 160343 175399: contig of 15057 bp in length
* 175400 175499: gap of 100 bp
* 175500 190856: contig of 15357 bp in length.

Location/Qualifiers

1. .190856
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1"
/clone="RP11-152L7"
/clone_lib="RPC1-11 Human Male BAC"
1. .1005
/note="assembly_fragment"
1106. .2321
/note="assembly_fragment"
2422. .4167
/note="assembly_fragment"
4268. .6402
/note="assembly_fragment"
6503. .9001
/note="assembly_fragment"
9102. .11665
/note="assembly_fragment"
11766. .13958
/note="assembly_fragment"
14059. .17065
/note="assembly_fragment"
17166. .20430
/note="assembly_fragment"
20531. .24620
/note="assembly_fragment"
24721. .28004
/note="assembly_fragment"
28105. .31520
/note="assembly_fragment"
31621. .35687
/note="assembly_fragment"
35788. .40442
/note="assembly_fragment"

```
misc_feature 40543..48006
              /note="assembly_fragment"
misc_feature 48107..55271
              /note="assembly_fragment"
misc_feature 55372..61182
              /note="assembly_fragment"
misc_feature 61283..68644
              /note="assembly_fragment"
misc_feature 68745..75906
              /note="assembly_fragment"
misc_feature 76007..83998
              /note="assembly_fragment"
misc_feature 84099..93489
              /note="assembly_fragment"
misc_feature 93590..101953
              /note="assembly_fragment"
misc_feature 102054..110821
              /note="assembly_fragment"
misc_feature 110922..121798
              /note="assembly_fragment"
misc_feature 121899..133647
              /note="assembly_fragment"
misc_feature 133748..146812
              /note="assembly_fragment"
misc_feature 146913..160242
              /note="assembly_fragment"
misc_feature 160343..175399
              /note="assembly_fragment"
misc_feature 175500..190856
              /note="assembly_fragment"
              clone_end:Sp6
              vector_side:right"
BASE COUNT 58822 a 35587 c 34792 g 58847 t 2808 others
ORIGIN
```

```
Query Match 4.7%; Score 22; DB 2; Length 190856;
Best Local Similarity 100.0%; Pred. No. 0.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gaaagagaaagaatgggaca 22
Db 21592 GAAAGAGAAAGAGATGGACA 21571
|||||
|||||
```

```
RESULT 9
LOCUS AL162294 61147 bp DNA PRI 01-MAR-2001
DEFINITION Human DNA sequence from clone RP5-852L11 on chromosome 20 Contains
ACCESSION AL162294
VERSION AL162294.8 GI:8218483
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 61147)
AUTHORS Lawlor,S.
TITLE Direct Submission
JOURNAL Submitted (26-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Jun 3, 2000 this sequence version replaced g1:8217681.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
```

on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr20>
IMPORTANT: This sequence is not the entire insert of clone RP5-852L11 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP4-800C24 is at 61048 in this sequence. The true right end of clone RP5-981M18 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP5-852L11 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pCYPAC2.

```
FEATURES
source 1..61147
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="20"
        /clone="RP5-852L11"
        /clone_lib="RPCI-5"
        complement(411..814)
        /note="match: GSS: Em:A0020910"
        611..972
        /note="MT1A1 repeat: matches 1..365 of consensus"
        1057..1793
        /note="L2 repeat: matches 117..959 of consensus"
        2150..2341
        /note="L2 repeat: matches 960..1150 of consensus"
        2500..2807
        /note="L2 repeat: matches 1152..1474 of consensus"
        2865..3227
        /note="L2 repeat: matches 1563..1966 of consensus"
        3569..4160
        /note="MER97a repeat: matches 2..605 of consensus"
        3752..4381
        /note="match: GSS: Em:A0051177"
        4378..4689
        /note="L1M7 repeat: matches 5823..6173 of consensus"
        5457..5504
        /note="24 copies 2 mer tc 95% conserved"
        5568..5867
        /note="Alusq repeat: matches 1..300 of consensus"
        5893..6432
        /note="MT2A repeat: matches 1..445 of consensus"
        6528..6816
        /note="Alusx repeat: matches 3..299 of consensus"
        8650..9161
        /note="MT2D repeat: matches 58..553 of consensus"
        9184..9625
        /note="LTR39 repeat: matches 301..745 of consensus"
        9626..10095
        /note="LTPA5 repeat: matches 5672..6141 of consensus"
        10096..10396
        /note="LTR39 repeat: matches 2..302 of consensus"
        11071..11241
        /note="FRAM repeat: matches 1..170 of consensus"
        11515..11818
        /note="Alu repeat: matches 1..310 of consensus"
        11839..11953
        /note="L2 repeat: matches 2630..2744 of consensus"
        12169..12788
        /note="L1MC4 repeat: matches 7055..7644 of consensus"
        12789..13090
        /note="Aluuo repeat: matches 1..299 of consensus"
```

repeat_region	13091. .13705
/note="L1MC4 repeat: matches 6464. .7055 of consensus"	
repeat_region	13795. .13906
/note="L2 repeat: matches 2629. .2741 of consensus"	
repeat_region	13934. .13969
/note="18 copies 2 mer tt 91% conserved"	
repeat_region	14474. .14969
/note="L2 repeat: matches 2241. .2750 of consensus"	
repeat_region	16139. .16434
/note="AluYb repeat: matches 1. .289 of consensus"	
repeat_region	17275. .17405
/note="L2 repeat: matches 2190. .2327 of consensus"	
repeat_region	17726. .18027
/note="AluSg repeat: matches 1. .309 of consensus"	
repeat_region	18576. .18669
/note="L1M3e repeat: matches 111. .213 of consensus"	
repeat_region	18670. .19124
/note="L1M3e repeat: matches 3. .449 of consensus"	
repeat_region	19266. .19424
/note="L2 repeat: matches 2344. .2504 of consensus"	
repeat_region	19479. .19559
/note="L2 repeat: matches 2626. .2720 of consensus"	
repeat_region	19712. .19814
/note="MER81 repeat: matches 1. .111 of consensus"	
repeat_region	20605. .20696
/note="46 copies 2 mer at 66% conserved"	
repeat_region	22494. .22795
/note="AluY repeat: matches 13. .303 of consensus"	
repeat_region	23166. .23241
/note="MSTD repeat: matches 1. .77 of consensus"	
repeat_region	23242. .23549
/note="AluY repeat: matches 1. .308 of consensus"	
repeat_region	23550. .23874
/note="MSTD repeat: matches 77. .394 of consensus"	
repeat_region	23945. .24183
/note="MIR repeat: matches 11. .252 of consensus"	
repeat_region	24726. .25020
/note="AluSg repeat: matches 1. .295 of consensus"	
repeat_region	25446. .25533
/note="L2 repeat: matches 490. .582 of consensus"	
repeat_region	25656. .25691
/note="MLT1J repeat: matches 112. .147 of consensus"	
repeat_region	25969. .26022
/note="17 copies 2 mer tt 91% conserved"	
repeat_region	26209. .26874
/note="L2 repeat: matches 803. .1526 of consensus"	
repeat_region	27517. .27802
/note="AluSx repeat: matches 1. .286 of consensus"	
repeat_region	27804. .27861
/note="29 copies 2 mer aa 69% conserved"	
repeat_region	27888. .28082
/note="L1ME3A repeat: matches 5683. .5863 of consensus"	
misc_feature	complement(27922. .28443)
repeat_region	/note="match: GSS: Em:AQ431424"
repeat_region	28435. .28598
/note="HERV1 repeat: matches 5609. .5761 of consensus"	
repeat_region	28602. .28916
/note="MLT2CB repeat: matches 9. .328 of consensus"	
repeat_region	29097. .29394
/note="AluSx repeat: matches 1. .299 of consensus"	
repeat_region	29795. .29851
/note="L1ME3A repeat: matches 5998. .6054 of consensus"	
repeat_region	30088. .30360
/note="L2 repeat: matches 2208. .2481 of consensus"	
repeat_region	33156. .34296
/note="L1MC4 repeat: matches 6755. .7928 of consensus"	
repeat_region	34545. .34709
/note="AluY repeat: matches 142. .306 of consensus"	
repeat_region	34977. .35252
/note="AluSx repeat: matches 44. .312 of consensus"	
misc_feature	35650. .36023
/note="match: GSS: Em:AQ116392"	
repeat_region	37255. .37554

```

/note="AluX repeat: matches 3. .302 of consensus"
38584. .38858
/note="THE1B repeat: matches 85. .364 of consensus"
39612. .39653
/note="21 copies 2 mer tc 81% conserved"
41265. .41480
/note="MIR repeat: matches 2. .236 of consensus"
41970. .42144
/note="MER5A repeat: matches 1. .189 of consensus"
42789. .42907
/note="L2 repeat: matches 2630. .2750 of consensus"
43218. .43340
/note="L2 repeat: matches 2623. .2750 of consensus"
43341. .43486
/note="L1MA8 repeat: matches 6130. .6284 of consensus"
43677. .44054
/note="L2 repeat: matches 2361. .2747 of consensus"
46456. .46688
/note="Alu repeat: matches 1. .302 of consensus"
47390. .47990
/note="MLT1F repeat: matches 1. .568 of consensus"
48024. .48331
/note="AluX repeat: matches 1. .304 of consensus"
48343. .48561
/note="MIR repeat: matches 30. .248 of consensus"
48560. .48676
/note="match: GSS: Em:AQ513873"
complement(48575. .48904)

```

Query Match	4.5%;	Score 21;	DB 9;	length 61147;
Best Local Similarity	100.0%;	Pred. No. 2.7;		
Matches	21;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;

QY 25 atgtgactgcctgatgaatt 45
|||||
Db 14001 ATGTGACTGCCTGATGAAGTT 13981

RESULT	10
AC026323	
LOCUS	
DEFINITION	AC026323 155190 bp DNA HTG 06-NOV-2000
ACCESSION	Homo sapiens chromosome 3 clone RP11-129P2, WORKING DRAFT SEQUENCE, 16 unordered pieces.
VERSION	AC026323
KEYWORDS	AC026323.9 GI:11094491
SOURCE	HTG; HTGS_PHASE1; HTGS_DRAFT. human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 155190)

Muzny,D.M., Adams,C., Adio-Oduola,B., All-ouman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsli,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lounseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,

Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Oguh,M., Okwuonu,G., Orangunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Slisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 155190)
Worley,K.C.
Direct Submission
Submitted (22-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 4, 2000 this sequence version replaced gi:8699803.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HAPV
Center clone name: RP11-129P2
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy: 12% of reads
Chemistry: Dye-terminator Big Dye: 88% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 136343 bases at least Q40
Consensus quality: 141554 bases at least Q30
Consensus quality: 144310 bases at least Q20
Estimated insert size: 144883; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 4.3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
43274: contig of 43274 bp in length
43374: gap of unknown length
43375: contig of 25135 bp in length
68510: gap of unknown length
68609: gap of unknown length
78881: contig of 10272 bp in length
78882: gap of unknown length
90715: contig of 11734 bp in length
90815: gap of unknown length
99704: contig of 8889 bp in length
99804: gap of unknown length
109357: contig of 9553 bp in length
109358: gap of unknown length
109458: gap of unknown length
118740: contig of 9283 bp in length
118840: gap of unknown length
126400: contig of 7560 bp in length
126500: gap of unknown length
134994: contig of 8494 bp in length

134995 135094: gap of unknown length
* 135095 139557: contig of 4463 bp in length
* 139558 139657: gap of unknown length
* 139658 143382: contig of 3725 bp in length
* 143383 143482: gap of unknown length
* 143483 147446: contig of 3964 bp in length
* 147447 147546: gap of unknown length
* 147547 150312: contig of 2766 bp in length
* 150313 150412: gap of unknown length
* 150413 152837: contig of 2425 bp in length
* 152838 152937: gap of unknown length
* 152938 154035: contig of 1098 bp in length
* 154036 154135: gap of unknown length
* 154136 155190: contig of 1055 bp in length.
Location/Qualifiers
1. 155190
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-129P2"
BASE COUNT 47836 a 29970 c 29096 g 46758 t 1530 others
ORIGIN
Query Match 4.5%; Score 21; DB 2; Length 155190;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 25 atgtgactgcctgatgaagt 45
Db 18469 ATGTGACTGCCTGATGAAGTT 18489
RESULT 11
AC005325
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCES
1 (bases 1 to 165228)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone hRPK.60_A_24
Unpublished
2 (bases 1 to 165228)
Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatlin,C.,
Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Hagos,B.,
Harris,K., Horton,L., Howland,J.C., Jacotot,L., Kann,L.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B., Peterson,K.,
Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
Strange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Torruella-Miller,I., Vassiliev,H., Vo,A.,
Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.
Direct Submission
Submitted (25-JUL-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 165228)
Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E., Devon,K.,
Dewar,K., Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B.,
Horton,L., Howland,J.C., Jacotot,L., Kann,L., Macdonald,P.,

TITLE
JOURNAL
COMMENT
Marquis,N., McEwan,P., McGurk,A., McKernan,K., Meldrim,J.,
Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nahf,R., Naylor,J.,
Niloff,M., O'Connor,T., Pavlin,B., Peterson,K., Riley,R.,
Roberts,D., Roy,A., Strange-Thomann,N., Stilwell,J., Stojanovic,N.,
Stone,C., Subramanian,A., Torrealia-Miller,I., Vassiliev,H., Vo,A.,
Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.
Direct Submission
Submitted (31-Jul-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 31, 1998 this sequence version replaced g1:3355498.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

FEATURES

source

Location/Qualifiers
1. 165228
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hRPK.60_A.24"
/clone_lib="RPCI-11 human BAC library"
/map="17"
/chromosome="17"
complement(351. 650)
/rpt_family="AluSg"
complement(986. 1310)
/rpt_family="MLT1A1"
complement(1636. 1775)
/rpt_family="AluJb"
complement(1770. 2048)
/rpt_family="AluSg"
complement(2055. 2329)
/rpt_family="L1ME"
complement(2597. 2813)
/rpt_family="MER58A"
complement(2936. 3055)
/rpt_family="FLAM_C"
complement(3366. 3610)
/rpt_family="MIR"
3624. 3676
/rpt_family="MLT1J"
3760. 3989
/rpt_family="MLT1J"
complement(5975. 6087)
/rpt_family="MLT1J"
8776. 9298
/rpt_family="MLT1H"
complement(9361. 9675)
/rpt_family="AluSg"
complement(9747. 10117)
/rpt_family="L1MB6"
10268. 10431
/rpt_family="MIR"
11205. 11253
/rpt_family="(CAT)n"
13213. 13294
/rpt_family="MIR"
13789. 13832
/rpt_family="(CA)n"
13833. 13867
/rpt_family="polypurine"
complement(14171. 14279)
/rpt_family="MIR"
14881. 14997
/rpt_family="MER33"
complement(14998. 15019)
/rpt_family="(GA)n"
complement(15027. 15182)
/rpt_family="AluJo"
complement(15239. 15303)
/rpt_family="AluJo"
15335. 15487
/rpt_family="MER33"
complement(16242. 16366)
/rpt_family="MIR"

repeat_region 16367. 16650
/rpt_family="AluY"
16651. 16688
/rpt_family="(CAAA)n"
complement(16689. 16731)
/rpt_family="MIR"
complement(16924. 16950)
/rpt_family="(CAGA)n"
complement(16951. 17159)
/rpt_family="AluSg"
complement(17166. 17436)
/rpt_family="AluSg"
19359. 19483
/rpt_family="MIR"
19484. 19785
/rpt_family="AluSx"
19786. 19840
/rpt_family="MIR"
complement(20472. 20642)
/rpt_family="MIR"
21141. 21424
/rpt_family="AluSx"
complement(21712. 21754)
/rpt_family="Alu"
complement(21798. 21946)
/rpt_family="MIR"
22298. 22395
/rpt_family="MIR"
complement(22543. 22841)
/rpt_family="AluSg"
24360. 24657
/rpt_family="AluSp"
25778. 25805
/rpt_family="AT-rich"
complement(25944. 26256)
/rpt_family="AluSp"
complement(26887. 27073)
/rpt_family="MER5A"
27095. 27143
/rpt_family="MER5A"
complement(27164. 27435)
/rpt_family="MLT1J"
27730. 27836
/rpt_family="MER81"
27837. 27967
/rpt_family="FLAM_C"
27968. 27982
/rpt_family="MER81"
27993. 28097
/rpt_family="MER5A"
28441. 28556
/rpt_family="L2"
28560. 28590
/rpt_family="(TGAA)n"
30757. 31192
/rpt_family="L2"
31193. 31351
/rpt_family="(TGGA)n"
31328. 31434
/rpt_family="(TGAA)n"
complement(31615. 31657)
/rpt_family="A-rich"
complement(31681. 31957)
/rpt_family="AluY"
32091. 32153
/rpt_family="MLT1J"
33172. 33226
/rpt_family="purine-rich"
complement(33256. 33303)
/rpt_family="L2"
complement(33508. 33610)
/rpt_family="L2"
complement(34078. 34245)
repeat_region

```
repeat_region      /rpt_family="MER63A"
                   complement(34259..34374)
repeat_region      /rpt_family="MIR"
                   complement(34446..34825)
repeat_region      /rpt_family="MSTD"
                   complement(34890..35015)
repeat_region      /rpt_family="L2"
                   complement(37119..37279)
repeat_region      /rpt_family="MIR"
                   complement(37786..38105)
repeat_region      /rpt_family="AluSx"
                   complement(38108..38393)
repeat_region      /rpt_family="AluSx"
                   complement(38574..38574)
repeat_region      /rpt_family="MSTD"
                   complement(38575..38873)

Query Match      4.5%; Score 21; DB 9; Length 165228;
Best Local Similarity 100.0%; Pred.No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      3 aagagaaagagaatggacag 23
          |||||
Db      88101 AAGAGAAAGAGATGGACAG 88121

RESULT 12
LOCUS   MTLPATP9S      1776 bp      DNA
DEFINITION L.perenne mitochondrial atp9 gene.
ACCESSION 250200
VERSION   Z50200.1 GI:1771354
KEYWORDS  atp9 gene; cytoplasmic-male-sterility; gene fusion.
SOURCE    Lolium perenne.
           Mitochondrion Lolium perenne
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
           Poideae; Poace; Lolium.
REFERENCE 1 (bases 1 to 1776)
AUTHORS   Kiang,S.A. and Kavanagh,T.T.
TITLE     Cytoplasmic male sterility (CMS) in Lolium perenne
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 1776)
AUTHORS   Kavanagh,T.T.
TITLE     Direct Submission
JOURNAL   Submitted (28-JUL-1995) Tony T.A. Kavanagh, Genetics, University of
           Dublin, Trinity College, Lincoln Place Gate, Dublin, Ireland
FEATURES
   source      1..1776
                /organism="Lolium perenne"
                /organelle="mitochondrion"
                /db_xref="taxon:4522"
   gene        651..875
                /gene="atp9"
   CDS         651..875
                /gene="atp9"
                /codon_start=1
                /protein_id="CAA90584.1"
                /db_xref="GI:1771355"
                /translation="MLEGAKLIGAGATIALAGAAIGIGNVSSLIHSVARNPSLAKQ
                LFGYAILGFALTETALALFALMMAFLLILEVF"

misc_feature    1239..1610
                /note="atp9 gene fusion"
BASE COUNT      523 a      325 c      419 g      509 t
ORIGIN
```

```
Query Match      4.3%; Score 20; DB 8; Length 1776;
Best Local Similarity 100.0%; Pred.No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      154 tttcagtggttttcctct 173
```

```
DB      1351 TTTCAAGTGTGTTCCTCT 1370
          |||||
RESULT 13
AF286474      2805 bp      DNA
LOCUS        AF286474
DEFINITION   Bos taurus retinitis pigmentosa GTPase regulator (RPGR) gene,
               partial cds.
ACCESSION    AF286474
VERSION      AF286474.1 GI:9837382
KEYWORDS     cow.
SOURCE       Bos taurus
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
               Bovidae; Bovinae; Bos.
REFERENCE    1 (bases 1 to 2805)
AUTHORS      Vervoort,R., Lennon,A., Bird,A.C., Tulloch,B., Axton,R.,
               Milano,M.G., Meindl,A., Meitinger,T., Ciccodicola,A. and Wright,A.F.
TITLE        Mutational hot spot within a new RPGR exon in X-linked retinitis
               pigmentosa
JOURNAL      Nat. Genet. (2000) In press
REFERENCE    2 (bases 1 to 2805)
AUTHORS      Vervoort,R., Lennon,A., Bird,A.C., Tulloch,B., Axton,R.,
               Milano,M.G., Meindl,A., Meitinger,T., Ciccodicola,A. and Wright,A.F.
TITLE        Direct Submission
JOURNAL      Submitted (13-JUL-2000) Cell Genetics, MRC Human Genetics Unit,
               Crewe Road, Edinburgh EH4 2XU, UK
FEATURES
   source      1..2805
                /organism="Bos taurus"
                /db_xref="taxon:9913"
   mRNA        <1..>2805
                /gene="RPGR"
                /product="retinitis pigmentosa GTPase regulator"
                /db_xref="GI:9837383"
   gene        <1..2805
                /gene="RPGR"
   CDS         <1..2805
                /gene="RPGR"
                /codon_start=1
                /product="retinitis pigmentosa GTPase regulator"
                /protein_id="AAG00553.1"
                /translation="KQELIEKPKQHTTYTESDSDSNYENNEISPKVTEGRVYKQLAQ
                GMYAMPVAISMETFSDEDVGDSDSQSPQSTSAEGLQKGFTRHKNKHDVYPLNTEEI
                EKESDEGQSQKSEANEIVSEMSDLVAKMTDLKDIRKTEENRKNTDFFDDLPRMYN
                IEDEEDKDFVCKSGDKODMIFDSEQASIEENSYLEGESSEQSIADGFQOSESTEF
                NSVEKDDDEVERTNQLMYSRFTFIQRHEARHRLSRIMAKYEFKCDRLSRIPBEOED
                SEGSGVEEOEIEANEECEGKEKEETALISDDLTDRAEVSEGKGLGGEAEHVPKGGG
                EGIRKEGNSGVEQSRSTGSEEGEEDMGECYIESLGKEKDLLEEEOQKEREQGHQ
                EERSTGRKEAGEROGGEGEGEEADDEEGEEDKEKEEGKEKEEEEOQEGEELIE
                GRWODGEEGEGKDEEGQEGGREGGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEG
                EEGGEEBAGEEGEAGEGEGKGREEEKGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEG
                EEEEGDGEEOQEAEGEGVEGDGEVEEVGDDEKEGDGEEEGDGEREGGTGEGDEEEOKG
                EEEEGDGEEGEEEGDGEEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEG
                EEEEGDEGEEEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEG
                EEEEGEELIEEGEGEGEGEAGEEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEG
                EGEGEREREGMEREEKEMVVRKRKEKENGRGEKEEDENNEEEKGKCOETGNEESGRQ
                GRKGGRESKVSКИRSGSKYKDRTHPKFTTNREGKGKGHGVRFKIPVQSKQLLVN
                GPPGSKKFWNNVLPHYLELK"

BASE COUNT      1089 a      207 c      1197 g      312 t
ORIGIN
```

```
QY      1 gaaagagaagaatggga 20
          |||||
Db      2463 GAAAGAGAAAGAGATGGGA 2482
```

RESULT 14
HSDJ130E4/c
LOCUS
DEFINITION HSDJ130E4 108893 bp DNA PRI 13-APR-2000
Human DNA sequence from clone RP1-130E4 on chromosome 6q24.2-25.3.
Contains the 3' end of the ESR1 gene for estrogen receptor 1, the
3' end of the gene KIAA0796, ESTs, STSS and GSSs, complete
sequence.
ACCESSION AL078582
VERSION AL078582.13 GI:7161747
KEYWORDS HTG; ESR1; estrogen receptor; KIAA0796.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 108893)
JOURNAL Parker,A.
Direct Submission
COMMENT Submitted (06-APR-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Mar 6, 2000 this sequence version replaced gi:7018350.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP1-130E4 is from the library RPCI-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone
RP1-130E4 It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP1-130E4 is at 108893 in this
sequence. The true right end of clone RP1-6315 is at 100 in this
sequence.
FEATURES
source location/Qualifiers
1..108893
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q24.2-25.3"
/clone="RP1-130E4"
/clone_lib="RPCI-1"
249..932
/note="match: GSS: Em:AQ373843"
1912..1967
repeat_region
/note="2 copies 28 mer 98% conserved"
2415..2462
repeat_region
/note="8 copies 6 mer caaaa 81% conserved"
2418..2463
repeat_region
/note="23 copies 2 mer aa 80% conserved"
3009..3407
repeat_region
/note="MSTB repeat: matches 1..426 of consensus"
3449..3681
repeat_region
/note="MER58A repeat: matches 1..214 of consensus"

repeat_region 4248..4299
/note="13 copies 4 mer ctcc 73% conserved"
repeat_region 4453..4742
/note="5 copies 58 mer 76% conserved"
repeat_region 4460..4733
/note="137 copies 2 mer tc 77% conserved"
repeat_region 4461..4732
/note="68 copies 4 mer ctct 77% conserved"
repeat_region 4463..4732
/note="45 copies 6 mer ctctct 78% conserved"
repeat_region 4467..4746
/note="10 copies 28 mer 76% conserved"
repeat_region 4476..4731
/note="2 copies 128 mer 82% conserved"
repeat_region 6026..6404
/note="L1MA6 repeat: matches 5497..5874 of consensus"
7053..7148
/note="MIR repeat: matches 29..124 of consensus"
8742..8789
/note="24 copies 2 mer ac 91% conserved"
9564..9615
/note="MIR repeat: matches 186..237 of consensus"
9743..9866
/note="L2 repeat: matches 2576..2709 of consensus"
10164..10258
/note="L1MB8 repeat: matches 6076..6171 of consensus"
10427..10669
/note="L1M4 repeat: matches 5115..5369 of consensus"
11871..12398
/note="match: GSS: Em:AQ439940"
11997..12046
/note="25 copies 2 mer tt 74% conserved"
join(<12052..12185,45446..45629,49793..54335)
/gene="ESR1"
/note="match: cDNAs: Em:AF124093 Em:AB007453 Em:M12674
Em:X03635 Em:AF042060 Em:Y00102 Em:AF042061 Em:U75604
Em:AF061275 Em:U75605 Em:AJ242740 Em:AJ242741 Em:AF119229
Em:AF095911 Em:X61098 Em:AF110402 Em:AB003356 Em:AB012721
Em:M38651 Em:Y18017 Em:AF099079 Em:AF136980 Em:Z37167
Em:AF172069 Em:AF172068 Em:AB036415 Em:AF128220
Em:AF128221 Em:AJ006039 Em:AF177465 Em:AF113513
Em:AF181077 Em:Z75126 Em:Z49257 Em:U47678 Em:M31559
Em:AF177936 Em:AF136979 Em:D28954 Em:X89959
match: ESTs: Em:AL040974 Em:A1634735 Em:A1758480
Em:A188116 Em:A1675862 Em:AW427925 Em:A1978964
Em:AA336802 Em:AW169860 Em:A1810941 Em:A1654044
Em:A1202659 Em:AA296578 Em:AA337245 Em:AA420328
Em:A1952266 Em:AV175085 Em:AW427924 Em:AW089672
Em:A1972926 Em:A1669577 Em:A1983175 Em:AA291702
Em:AW088459 Em:AW438839 Em:AA319335 Em:AA164585
Em:A1654018 Em:AA164586 Em:A1625626 Em:T28397 Em:A1673252
Em:AA336938 Em:AA336971 Em:A1540656 Em:AA299347
Em:AA362012 Em:AL040975 Em:AU018232 Em:AW440601
Em:A1669190 Em:AA337706 Em:AA331020 Em:A1888281 Em:C02879
Em:A1984079"
/evidence=not_experimental
/product="dJ130E4.1 (estrogen receptor 1)"
12052..54335
/gene="ESR1"
join(<12052..12185,45446..45629,49793..50027)
/gene="ESR1"
/note="match: proteins: Sw:P03372 Tr:Q95131 Sw:P81559
Sw:P06211 Sw:P06212 Sw:P19785 Sw:Q91250 Tr:Q9YHT3
Sw:Q29040"
/codon_start=2
/evidence=not_experimental
/product="dJ130E4.1 (estrogen receptor 1)"
/protein_id="CAB87585.1"
/db_xref="GI:7573520"
/translation="NQGKVEGAVEIFDMLATSSRFPMNLQGEFVCLKSTILNS
GVYFLSTPLKSLSEKDHIRVLDKITDTLLHLMKAGLTLQQOQRLAQLLILSHI
RHMSNKGMEHLYSMKCKNVVPLVDLLEMLDAHRLHAPSRGASVEETDQSHLATAG
STSSHSLOKYYITGEAGEGPATV"


```
repeat_region      12791..12834
                    /note="22 copies 2 mer tt 81% conserved"
repeat_region      12796..12831
                    /note="6 copies 6 mer tttttg 88% conserved"
repeat_region      13192..13447
                    /note="2 copies 128 mer 92% conserved"
repeat_region      14237..14286
                    /note="2 copies 25 mer 96% conserved"
repeat_region      15390..15551
                    /note="L2 repeat: matches 2364..2537 of consensus"
repeat_region      15599..15658
                    /note="L2 repeat: matches 2702..2750 of consensus"
misc_feature       complement(17729..18056)
                    /note="match: GSS: Em:AQ061898"
repeat_region      17821..18050
                    /note="MIR repeat: matches 23..252 of consensus"
misc_feature       17877..18256
                    /gene="ESR1"
                    /note="match: GSS: Em:AQ052854"
repeat_region      18641..18916
                    /note="AluSq repeat: matches 37..306 of consensus"
repeat_region      19498..19640
                    /note="MER5B repeat: matches 1..140 of consensus"
repeat_region      20105..20541
                    /note="LTR26 repeat: matches 151..601 of consensus"
repeat_region      20573..20927
                    /note="MER1B repeat: matches 1..337 of consensus"
repeat_region      22022..22071
                    /note="MIR repeat: matches 208..257 of consensus"
repeat_region      22029..22082
                    /note="L2 repeat: matches 2639..2692 of consensus"
repeat_region      22087..22152
                    /note="MIR repeat: matches 116..182 of consensus"
repeat_region      22449..22639
                    /note="MIR repeat: matches 8..224 of consensus"
repeat_region      24767..24850
                    /note="MIR repeat: matches 22..100 of consensus"
repeat_region      25014..26152
                    /note="L1MC1 repeat: matches 5162..6324 of consensus"
misc_feature       26439..26977
                    /gene="ESR1"
                    /note="match: GSS: Em:AQ677837"
misc_feature       26664..27171
                    /gene="ESR1"
                    /note="match: GSS: Em:AQ081793
                    match: STS: Em:G53308"
misc_feature       26665..27078
                    /gene="ESR1"
                    /note="match: GSS: Em:AQ825841"
repeat_region      27152..27515
                    /note="THE1A repeat: matches 1..354 of consensus"
repeat_region      27516..29089
                    /note="THE1A-internal repeat: matches 1..1580 of
                    consensus"
repeat_region      29090..29326
                    /note="THE1A repeat: matches 1..238 of consensus"
repeat_region      29566..34858
```

```
Query Match      4.3%; Score 20; DB 9; Length 108893;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 gaaagagaagaagaatggga 20
        |||
Db      4581 GAAAGAGAAAGAGATGGGA 4562
```

```
RESULT 15
PFMAL4P1_0
WPCOMMENT
Sequence split into 4 fragments LOCUS PFMAL4P1 Accession AL034557
Fragment Name Begin End
PFMAL4P1_0 1 11000
```

```
PFMAL4P1_1      100001      210000
PFMAL4P1_2      200001      310000
PFMAL4P1_3      300001      392633
LOCUS      PFMAL4P1 392633 bp DNA HTG 11-AUG-1999
DEFINITION      Plasmodium falciparum chromosome 4 strain 3D7, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION      AL034557
VERSION      AL034557.7 GI:5731897
KEYWORDS      HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
ORGANISM      Plasmodium falciparum
REFERENCE      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS      1 (bases 1 to 392633)
              Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.
              and Barrell,B.
TITLE      Direct Submission
JOURNAL      Submitted (17-DEC-1998) P.falciparum Genome Sequencing Consortium,
              The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
              CB10 1SA, UK
COMMENT      On Aug 12, 1999 this sequence version replaced gi:5531346.
              For more information about this sequence or the Malaria Project,
              see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
              sequence is unfinished and does not necessarily represent the
              correct sequence. Work on the sequence is in progress and the
              release of this data is based on the understanding that the
              sequence may change as work continues. The sequence may be
              contaminated with foreign sequence from E.coli, yeast, vector,
              phage etc.
              Order of segments is not known; 800 n's separate segments.
              * NOTE: This is a 'working draft' sequence.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
FEATURES
    source          location/Qualifiers
                    1..392633
                    /organism="Plasmodium falciparum"
                    /strain="3D7"
                    /db_xref="taxon:5833"
                    /chromosome="4"
BASE COUNT      146453 a 41490 c 42490 g 138992 t 23208 others
ORIGIN
```

```
Query Match      4.3%; Score 20; DB 2; Length 110000;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 gaaagagaagaagaatggga 20
        |||
Db      30271 GAAAGAGAAAGAGATGGGA 30290
```

Search completed: April 6, 2002, 06:18:43
Job time: 21950 sec

Sat Apr 6 11:03:26 2002

us-09-621-781-1_copy_1507_1970.rge

Page 14

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2002, 04:29:53 ; Search time 157.36 Seconds
(without alignments)
2527.954 Million cell updates/sec

Title: US-09-621-781-1_COPY_1507_1970
Perfect score: 464
Sequence: 1 gaaagagaagaatggga.....tttgcaaccgacggac 464

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_1101:*

1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	223	48.1	2614	19	AAV65766	Rat progression el
2	223	48.1	2614	20	AAZ23030	Rat progression el
3	108	23.3	2111	19	AAV65765	Human progression
4	108	23.3	2111	20	AAZ23029	Human progression
5	108	23.3	2111	20	AAZ23029	Human progression
6	19	4.1	72928	20	AAZ18355	Human ASTH1J 5' ge
7	19	4.1	72928	21	AAA80253	Human ASTH1J 5' ge
8	18	3.9	819	22	AAH05417	Human CDNA clone (
9	17	3.7	686	19	AAZ14334	H. pylori GHPO 212
10	17	3.7	1248	21	AAA26436	Human secreted pro
11	17	3.7	1278	21	AAZ76612	Human ORFX ORF2167

12	17	3.7	1648	22	AAC87123	Nucleotide sequenc
13	17	3.7	1775	22	AAC87122	Nucleotide sequenc
14	17	3.7	2075	21	AAZ44199	Arabidopsis thalia
15	17	3.7	3122	16	AAZ79913	Lys-aminopeptidase
16	17	3.7	3591	22	AAH16646	Human CDNA sequenc
17	17	3.7	3729	22	AAZ77688	Human wild-type Fc
18	17	3.7	4131	22	AAZ77689	Human variant Fcep
19	17	3.7	5538	22	AAH18660	Human CDNA sequenc
20	17	3.7	11298	18	AAZ86756	Human high affinity
21	17	3.7	11298	19	AAV54661	Human high affinity
22	17	3.7	11298	21	AAZ20937	Human high affinity
23	17	3.7	11298	21	AAA34815	Human adenosine re
24	17	3.7	11298	22	AAZ92144	Human IGERB gene S
25	17	3.7	11357	14	AAZ51024	Human FCER1 beta C
26	17	3.7	21742	21	AAZ20938	Human high affinity
27	17	3.7	21742	21	AAZ20938	Human high affinity
28	17	3.7	28720	19	AAV49655	Human adenosine re
29	17	3.7	33030	22	AAZ29337	Human IGERB gene S
30	17	3.7	117609	21	AAZ21435	Human SC3 DNA. Ho
31	16	3.4	63	21	AAZ11216	Atopy related gene
32	16	3.4	114	22	AAZ146006	Human secreted pro
33	16	3.4	161	21	AAZ26461	Probe #14692 used
34	16	3.4	232	14	AAZ39840	Human prostate can
35	16	3.4	232	14	AAZ59252	Expressed Sequence
36	16	3.4	253	21	AAA26462	Human prostate can
37	16	3.4	287	21	AAZ09789	Human brain Expres
38	16	3.4	296	22	AAZ57333	Human brain specif
39	16	3.4	340	22	AAZ122552	Human brain specif
40	16	3.4	340	22	AAZ17842	Probe #12485 for g
41	16	3.4	340	22	AAZ108235	Probe #16528 used
42	16	3.4	341	22	AAZ91993	B thuringiensis 14
43	16	3.4	389	22	AAZ57541	Human brain cell s
44	16	3.4	417	20	AAZ8497	EST clone GT6. Ho
45	16	3.4	446	21	AAZ29007	Human secreted pro

ALIGNMENTS

RESULT 1	
AAV65766	standard; CDNA; 2614 BP.
ID	AAV65766
XX	AAV65766;
AC	
XX	
DT	02-FEB-1999 (first entry)
XX	
DE	Rat progression elevated gene-2 (PEG-3) promoter region.
XX	
KW	Progression elevated gene-3; PEG-3; rat; tumour progression;
KW	DNA repair; tumorigenesis; angiogenesis; metastasis; melanoma;
KW	brain cancer; cervix cancer; prostate cancer; lung cancer;
KW	colorectal cancer; neuroblastoma; glioblastoma; diagnosis; therapy;
KW	promoter; ss.
XX	
OS	Rattus sp.
XX	
FH	
FT	Key
FT	misc_feature
FT	Location/Qualifiers
FT	128..133
FT	/*tag= a
FT	/label= PEA3
FT	2128..2133
FT	/*tag= b
FT	/label= PEA3
FT	2404..2409
FT	/*tag= c
FT	/label= PEA3
FT	338..344
FT	/*tag= d
FT	/label= E2A
FT	351..357
FT	/*tag= e
FT	/label= GRE
FT	

FT misc_feature 1094..1099
FT /tag= f
FT /label= GRE
FT misc_feature 729..737
FT /tag= g
FT /label= E2F
FT misc_feature 193..197
FT /tag= h
FT /label= TRE
FT misc_feature 489..493
FT /tag= i
FT /label= TRE
FT misc_feature 936..940
FT /tag= j
FT /label= TRE
FT misc_feature 774..779
FT /tag= k
FT /note= "acute phase reactive regulating element"
FT 2090..2095
FT /tag= l
FT /note= "acute phase reactive regulating element"
FT 2028..2033
FT /tag= m
FT /label= SP1
FT misc_feature 1784..1790
FT /tag= n
FT /label= AP1
FT misc_feature 2552..2558
FT /tag= o
FT /label= AP1
FT misc_feature 865..874
FT /tag= p
FT /label= AP2
FT misc_feature 1597..1609
FT /tag= q
FT /label= NF1
XX
PN WC9842315-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US05793.
XX
PR 21-MAR-1997; 97US-0821818.
XX
PA (GENO-) GENQUEST INC.
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI Fisher PB, Goldstein NI, Su Z, Zhang N;
XX
DR WPI; 1998-557025/47.
XX
PT New isolated Progression Elevated Gene-3 - used to develop products
PT for e.g. modulating DNA damage and repair pathways, cancer
PT progression or oncogene mediated transformation and angiogenesis.
XX
PS Claim 24; Fig 14A-B; 225pp; English.
XX
CC This is the 5' regulatory region of the rat progression elevated
CC gene-3 (PEG-3, see also AAV65764). It was identified using a genomic
CC walking strategy, and appears to encompass a functionally complete
CC PEG-3 gene promoter including a number of potentially important
CC transcriptional motifs. PEG-3 expression is uniquely elevated in
CC all cases of rodent progression analysed, and the level of
CC expression is also an indicator of DNA damage in that cell. Cells
CC in which a reporter gene is under control of the promoter of the
CC PEG-3 gene can be used in methods for identifying agents that
CC modulate PEG-3 expression or the ability of PEG-3 to induce
CC progression, or for determining whether an agent is capable of
CC inhibiting DNA damage and repair pathways, cancer progression or
CC oncogene-mediated transformation. Compounds that induce DNA
CC damage or which regulate angiogenesis can also be identified using
CC such cells. The methods can be applied to a progression phenotype

CC comprising anchorage-independent growth, tumorigenesis, angiogenesis
CC or metastasis, to melanoma, brain, cervical, prostate, lung or
CC colorectal cancer, neuroblastoma or glioblastoma.
XX
SQ Sequence 2614 BP; 737 A; 615 C; 706 G; 556 T; 0 other;

Query Match 48.1%; Score 223; DB 19; Length 2614;
Best Local Similarity 100.0%; Pred. No. 8.5e-110;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaagagaagaagatgggacagcatgtgactgcctgatgaagttggcgtgtgtctcaa 60
Db 1512 gaaagagaagaagatgggacagcatgtgactgcctgatgaagttggcgtgtgtctcaa 1571
QY 61 aagttctgcgagattgacggtctctgtgattgaagccaagacacgctgggaagccag 120
Db 1572 aagttctgcgagattgacggtctctgtgattgaagccaagacacgctgggaagccag 1631
QY 121 gtgacctcacaaagcccggaatctccgcgagaatttcagtgttttcctctccacc 180
Db 1632 gtgacctcacaaagcccggaatctccgcgagaatttcagtgttttcctctccacc 1691
QY 181 ttctcagggacttcggaactccgcctctccggtgacgtcag 223
Db 1692 ttctcagggacttcggaactccgcctctccggtgacgtcag 1734

RESULT 2
AAZ23030
ID AAZ23030 standard; DNA; 2614 BP.
XX
AC AAZ23030;
XX
DT 17-JAN-2000 (first entry)
XX
DE Rat progression elevated gene-3 (PEG-3) promoter sequence.
XX
KW Progression-elevated gene-3; PEG-3 gene; regulatory region; cell death;
KW cell growth; cancer; thymidine kinase; gancyclovir; acyclovir; rat;
KW promoter; ss.
XX
OS Rattus sp.
XX
PN WO9949898-A1.
XX
PD 07-OCT-1999.
XX
PF 31-MAR-1999; 99WO-US07199.
XX
PR 31-MAR-1998; 98US-0052753.
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI Fisher PB;
XX
DR WPI; 1999-591184/50.
XX
PT Novel vectors useful for studying the progression of cancer -
XX
PS Disclosure; Fig 14 A-B; 251pp; English.
XX
CC The invention relates to an inducible progression-elevated gene-3 (PEG-3
CC gene) regulatory region functionally linked to a gene encoding a product
CC that causes or may be induced to cause the death or inhibition of cancer
CC cell growth. A vector of the invention which contains a gene encoding
CC thymidine kinase or a product which causes the cell to express a
CC specific antigen can be administered along with gancyclovir or acyclovir,
CC or an antibody or fragment to the antigen, respectively, to treat cancer
CC in a subject. The PEG-3 gene is useful for generating new cloning and
CC expression vectors, transfected cells, and for developing methods for
CC cultured growth of such cells. The PEG-3 polynucleotide is also useful as
CC a source of primers and probes to study the progression of cancer, and to

CC detect the presence of the gene. The present sequence represents the
CC nucleotide sequence of the rat PEG-3 gene promoter.
XX
SQ Sequence 2614 BP; 737 A; 615 C; 706 G; 556 T; 0 other;

Query Match 48.1%; Score 223; DB 20; Length 2614;
Best Local Similarity 100.0%; Pred. No. 8.5e-110;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gaaagagaaagagaatgagacagcatgtgactgcctgataaagttggcgtgtgtctcaaa 60
Db 1512 gaaagagaaagagaatgagacagcatgtgactgcctgataaagttggcgtgtgtctcaaa 1571
OY 61 aagttctgcgagattgacgctctcttgatttgagccaagacacgcctgggaagccacg 120
Db 1572 aagttctgcgagattgacgctctcttgatttgagccaagacacgcctgggaagccacg 1631
OY 121 gtgacctacacaagcccggaatctccgcgagaatttcagttgttttcctctctccacc 180
Db 1632 gtgacctacacaagcccggaatctccgcgagaatttcagttgttttcctctctccacc 1691
OY 181 ttcttcaggacttcggaactccgcctctccggtgacgtcag 223
Db 1692 ttcttcaggacttcggaactccgcctctccggtgacgtcag 1734

RESULT 3
AAV65765
ID AAV65765 standard; cDNA; 2111 BP.

AC AAV65765;
DT 02-FEB-1999 (first entry)

DE Human progression elevated gene-2 (PEG-3) cDNA.

KW Progression elevated gene-3; PEG-3; human; tumour progression;
KW DNA repair; tumorigenesis; angiogenesis; metastasis; melanoma;
KW brain cancer; cervix cancer; prostate cancer; lung cancer;
KW colorectal cancer; neuroblastoma; glioblastoma; diagnosis; therapy;
ss.

OS Homo sapiens.

FH Key Location/Qualifiers
FT CDS 294..2030
FT /*tag= a

PN WO9842315-A1.

PD 01-OCT-1998.

PF 20-MAR-1998; 98WO-US05793.

PR 21-MAR-1997; 97US-0821818.

PA (GENO-) GENQUEST INC.
(UYCO) UNIV COLUMBIA NEW YORK.

PI Fisher PB, Goldstein NI, Su Z, Zhang N;

DR WPI; 1998-557025/47.

DR P-PSDB; AAW79958.

PT New isolated Progression Elevated Gene-3 - used to develop products
PT for e.g. modulating DNA damage and repair pathways, cancer
PT progression or oncogene mediated transformation and angiogenesis.

PS Claim 4; Fig 13A-C; 225pp; English.

CC This nucleotide sequence includes an open reading frame encoding
CC human progression elevated gene-3 (PEG-3) protein (see AAW79957).

CC PEG-3 clones were isolated from an MCF-7 cDNA library using rat
CC PEG-3 cDNA (see AAV65764) as probe. 5'RACE (see AAV65767-68) was used
CC to generate full-length cDNA. PEG-3 mRNA is highly expressed in
CC most human tumour cell lines. Probes derived from the gene can be
CC used to monitor tumour progression. Antisense sequences can be
CC used to inhibit tumour progression. Cells in which a reporter gene
CC is under control of the promoter (see AAV65766) of the rat PEG-3 gene
CC can be used in methods for identifying agents that modulate PEG-3
CC expression or the ability of PEG-3 to induce progression, or for
CC determining whether an agent is capable of inhibiting DNA damage
CC and repair pathways, cancer progression or oncogene-mediated
CC transformation. Compounds that induce DNA damage or which regulate
CC angiogenesis can also be identified using such cells. Transgenic
CC animals and vaccines comprising PEG-3 polypeptides and an immune
CC response enhancer are also claimed. The methods can be applied to
CC a progression phenotype comprising anchorage-independent growth,
CC tumorigenesis, angiogenesis or metastasis, to melanoma, brain,
CC cervical, prostate, lung or colorectal cancer, neuroblastoma or
CC glioblastoma.

SQ Sequence 2111 BP; 533 A; 607 C; 575 G; 396 T; 0 other;

Query Match 23.3%; Score 108; DB 19; Length 2111;
Best Local Similarity 100.0%; Pred. No. 4.2e-48;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 332 cctttccgagacagcctttgcagacagcccgtagacatcacgtcccgagcccccacgcc 391
Db 120 cctttccgagacagcctttgcagacagcccgtagacatcacgtcccgagcccccacgcc 179

OY 392 tgaaggcgacatgaacgcgctgacctgagagacaatccgaccacga 439
Db 180 tgaaggcgacatgaacgcgctgacctgagagacaatccgaccacga 227

RESULT 4
AAZ23029
ID AAZ23029 standard; cDNA; 2111 BP.
AC AAZ23029;

DT 17-JAN-2000 (first entry)

DE Human progression elevated gene-3 (PEG-3) cDNA sequence.

XX Progression-elevated gene-3; PEG-3 gene; regulatory region; cell death;
KW cell growth; cancer; thymidine kinase; gancyclovir; acyclovir; human; ss.

OS Homo sapiens.

PN WO9949898-A1.

PD 07-OCT-1999.

PF 31-MAR-1999; 99WO-US07199.

PR 31-MAR-1998; 98US-0052753.

PA (UYCO) UNIV COLUMBIA NEW YORK.

PI Fisher PB;

DR WPI; 1999-591184/50.

DR P-PSDB; AAY41104.

PT Novel vectors useful for studying the progression of cancer -
PT Disclosure; Fig 13A-C; 251pp; English.

CC The invention relates to an inducible progression-elevated gene-3 (PEG-3
CC gene) regulatory region functionally linked to a gene encoding a product
CC that causes or may be induced to cause the death or inhibition of cancer

CC cell growth. A vector of the invention which contains a gene encoding
CC thymidine kinase or a product which causes the cell to express a
CC specific antigen can be administered along with gancyclovir or acyclovir,
CC or an antibody or fragment to the antigen, respectively, to treat cancer
CC in a subject. The PEG-3 gene is useful for generating new cloning and
CC expression vectors, transfected cells, and for developing methods for
CC cultured growth of such cells. The PEG-3 polynucleotide is also useful as
CC a source of primers and probes to study the progression of cancer, and to
CC detect the presence of the gene. The present sequence represents the
CC nucleotide sequence of the human PEG-3 cDNA.
XX
SQ Sequence 2111 BP; 533 A; 608 C; 575 G; 395 T; 0 other;

Query Match 23.3%; Score 108; DB 20; Length 2111;
Best Local Similarity 100.0%; Pred. No. 4.2e-48;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 cctttgccgggacagcctttgagacagcccgtagacatcacgtcccgagccccacgcc 391
b ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
120 ccttgccgggacagcctttgagacagcccgtagacatcacgtcccgagccccacgcc 179
QY 392 tgaggggcagcatgaacgcgtgtgaccttgagagcaatccggaaccacga 439
b ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 180 tgaggggcagcatgaacgcgtgtgaccttgagagcaatccggaaccacga 227

RESULT 5

AAx87383 ID AAX87383 standard; cDNA; 2111 BP.

XX AC AAX87383;

DT 08-OCT-1999 (first entry)

DE Human progression elevated gene-3 (PEG-3) cDNA.

KW Pifogression elevated gene-3; PEG-3; human; angiogenesis; cancer;
therapy; vaccine; ss.

OS Homo sapiens.

FT Key Location/Qualifiers
CDS 294..2030
/*tag= a

PN WO9937776-A1.

PD 29-JUL-1999.

PF 26-JAN-1999; 99WO-US01623.

PR 29-JAN-1998; 98US-0072941.
26-JAN-1998; 98US-0072469.

PA (GENQ-) GENQUEST INC.

PI Fisher PB, Zhang N;

DR WPI: 1999-458694/38.
P-PSDB; AAY06514.

PT Modulation of angiogenesis by altering the expression and/or
PT activity of a progression-associated protein, especially for cancer
PT treatment

PS Claim 1; Page 71-74; 81pp; English.

CC This is the nucleotide sequence of a cDNA clone corresponding to
CC the human progression elevated gene-3 (PEG-3). It codes for a
CC 578-amino acid protein (see AAY06514), and represents a new member
CC of the gadd34/MyD16 gene family. The human PEG-3 cDNA was
CC isolated from a MCF-7 cDNA library using rat PEG-3 (see AAX87387) as

CC probe. Human PEG-3 is generally expressed in cells that are in
CC progression, including most human tumour cell lines. A claimed
CC method for modulating angiogenesis in an organism comprises
CC administering an agent that alters expression and/or activity of a
CC PEG-3 protein, such as an antisense polynucleotide or antibody.
CC Also claimed is a method for determining whether an agent
CC modulates angiogenesis, where the candidate agent is present
CC within a combinatorial small molecule library. Vaccines and
CC pharmaceutical compositions comprising such compounds are also
CC provided and may be used to prevent angiogenesis, especially
CC related to cancer cell progression.
XX

SQ Sequence 2111 BP; 533 A; 608 C; 575 G; 395 T; 0 other;

Query Match 23.3%; Score 108; DB 20; Length 2111;
Best Local Similarity 100.0%; Pred. No. 4.2e-48;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 cctttgccgggacagcctttgagacagcccgtagacatcacgtcccgagccccacgcc 391
b ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 120 ccttgccgggacagcctttgagacagcccgtagacatcacgtcccgagccccacgcc 179
QY 392 tgaggggcagcatgaacgcgtgtgaccttgagagcaatccggaaccacga 439
b ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 180 tgaggggcagcatgaacgcgtgtgaccttgagagcaatccggaaccacga 227

RESULT 6

AAZ18355 ID AAZ18355 standard; DNA; 72928 BP.

XX AC AAZ18355;

DT 19-OCT-1999 (first entry)

DE Human ASTH1J 5' genomic region.

KW ASTH1; asthma; human; chromosome 11p; ASTH1I; ASTH1J; genetic locus;
therapeutic; immunogen; ds.

OS Homo sapiens.

PN WO9937809-A1.

PD 29-JUL-1999.

PF 21-JAN-1998; 98WO-US01260.

PR 21-JAN-1998; 98WO-US01260.

PA (AXYS-) AXYS PHARM INC.

PI Brooks-Wilson AR, Buckler A, Cardon L, Carey AH;
PI Galvin M, Miller A, North M;

DR WPI: 1999-479058/40.

PT Mammalian asthma related genes, useful for diagnosis of a
PT predisposition to development of asthma

PS Claim 15; Page 75-96; 195pp; English.

CC The invention identifies a genetic locus ASTH1, associated with asthma,
CC mapped to human chromosome 11p. ASTH1I and ASTH1J are genes present
CC within the locus, located close to each other on human chromosome 11p,
CC and have similar patterns of expression, and common sequence motifs. The
CC ASTH1 genes and fragments, encoded protein, genomic regulatory regions
CC and anti-ASTH1 antibodies are useful in the identification of individuals
CC predisposed to development of asthma, and for the modulation of gene
CC activity in vivo for prophylactic and therapeutic purposes. The ASTH1
CC protein is useful as an immunogen to raise specific antibodies, in drug
CC screening for compositions that mimic or modulate ASTH1 activity or

CC expression, including altered forms of ASTH1 protein, and as a
CC therapeutic. The present sequence represents a human ASTH1 genomic region
CC sequence.

Sequence 72928 BP; 20042 A; 15100 C; 16036 G; 21750 T; 0 other;

Query Match	4.1%;	Score 19;	DB 20;	Length 72928;
Best Local Similarity	100.0%;	Pred. No. 2.4;		
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

```
QY      2 aaagagaaagagaatgga 20
         |||||
Db 46222 aaagagaaagagaatgga 46240
```

RESULT	7
AAA80253	
ID	AAA80253 standard; DNA; 72928 BP.
XX	
AC	AAA80253;
XX	
DT	22-NOV-2000 (first entry)
XX	
DE	Human ASTH1J 5' genomic region.

KW ASTH1 locus; ASTH1J; human; chromosome 11p; asthma;
KW bronchial hyperreactivity; ets family; transcription factor;
KW splice variant; genetic predisposition; polymorphism; antibody;;
KW drug screening; prophylaxis; therapy; diagnosis; ds.

OS Homo sapiens.

PN US6087485-A.

PD 11-JUL-2000.

PF 21-JAN-1998; 98US-0009913.

PR 21-JAN-1997; 97US-0035663.

PR 01-JUL-1997; 97US-0051432.

PA (AXYS-) AXYS PHARM INC.

Galvin M, Miller A, North M, Cardon L, Buckler A;

PI Brooks-Wilson AR, Carey AH;

DR WPI; 2000-505109/45.

PT	New nucleic acids other than naturally occurring chromosomes encoding
PT	ASTH1 protein, for e.g. screening compositions that modulate expression
PT	or function of ASTH1 proteins or as diagnostics for genetic
PT	predisposition to asthma -

PS Claim 7; Column 49-112; 131pp; English.

The invention relates to the ASTH1 locus on the short arm of human chromosome (11p). This locus comprises the ASTH1 and ASTH1J genes, which are associated with a genetic predisposition to asthma and bronchial hyperactivity. The ASTH1 and ASTH1J genes are oriented in opposite directions with the ASTH1 locus, and have similar patterns of expression and common sequence motifs. They are both expressed in trachea, lung and several other tissues. ASTH1 and ASTH1J are novel members of the ets family of transcription factors, which have been implicated in the activation of a variety of genes including the TCRA gene and cytokine genes known to be important in the aetiology of asthma. Both ASTH1 and ASTH1J mRNAs are alternatively spliced. Alternative splicing of transcripts has no effect on the open reading frame of ASTH1J, as the exons involved are all 5' to the start codon in exon b. In contrast, alternative splicing of ASTH1 transcripts results in 3 different ASTH1 isoforms. The invention also encompasses mouse asth1j protein. The ASTH1 nucleic acids are useful as diagnostics to identify a hereditary predisposition to asthma, as probes for identifying ASTH1

CC related genes, for identifying expression of the gene in a biological
CC specimen, and for generating genetically modified non-human animals or
CC site specific gene modifications in cell lines. The encoded ASTH1
CC proteins are useful as immunogens to raise specific antibodies; in drug
CC screening for compositions that mimic or modulate activity or expression
CC of ASTH1 and/or ASTHJ (including altered forms of these proteins); and
CC as a therapeutic. The ASTH1 genes or fragments thereof, encoded proteins,
CC ASTH1 genomic regulatory regions, and anti-ASTH1 and anti-ASTHJ
CC antibodies are useful in the identification of individuals predisposed to
CC development of asthma, and for modulation of gene activity in vivo for
CC prophylactic and therapeutic purposes. The intact ASTH1 or ASTHJ
CC proteins or active fragments thereof may be used to modulate or reduce
CC bronchial hyperreactivity. The present sequence represents the
CC non-transcribed region upstream of the human ASTHJ gene.

Sequence 72928 BP; 20041 A; 15101 C; 16036 G; 21750 T; 0 other; 50

Query Match	4.1%;	Score 19;	DB 21;	Length 72928;
Best Local Similarity	100.0%;	Pred. No. 2.4;		
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 2 aaagaagaaagaaatgga 20
|||||
Db 46222 aaagaagaaagaaatgga 46240

RESULT	8
AAH05417/c	
ID	AAH05417 standard; cDNA; 819 bp.

AC AAH05417;

DT 26-JUN-2001 (first entry)

AA	DE	Human cDNA clone (5'-primer)	SEQ ID NO:2252.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

05 Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

AA 28-JUL-2000; 2000EP-0116126.
PF

29-JUL-1999: 99JP-0248036.

PR 11-JAN-2000: 2000JP-0118776.

PR 09-JUN-2000: 2000JP-0241899

XX
PA (HELI-) HELIX RES INST.

XX
PT Ota T., Tsogai T., Nishikawa T., Hayashi K., Saito K., Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
PI Oda I, Isogai I, Nishikawa A, Nagasawa K, Otsuki T;

AA
DB
WPT: 2001-318749/34.

aa primer sets for synthesizing polynucleotides, particularly the 5602
 pt full-length cDNAs defined in the specification, and for the detection
 pt and/or diagnosis of the abnormality of the proteins encoded by the
 pt full-length cDNAs -

aa	
PS	Claim 1; SEQ ID 2252; 2537bp + CD ROM; English.

The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

SQ Sequence 819 BP; 171 A; 286 C; 156 G; 203 T; 3 other;

Query Match 3.9%; Score 18; DB 22; Length 819;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 agagaatggagacgcatg 27
|||||
Db 786 AGAGAATGGAGACGACATG 769

RESULT 9
AAAX14334
ID AAAX14334 standard; DNA; 686 BP.
XX
AC AAAX14334;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 212 gene.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 47..631
FT /*tag= a
XX
PN WO9843478-A1.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX
DR WPI; 1998-542293/46.
DR P-PSDB; AAW98615.
XX
XX New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
PS Claim 1; Page 1242-1243; 2054pp; English.
XX
CC This sequence represents a polynucleotide of the invention. It was

CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.

SQ Sequence 686 BP; 221 A; 123 C; 161 G; 181 T; 0 other;

Query Match 3.7%; Score 17; DB 19; Length 686;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 tgctgtatgaattgac 48
|||||
Db 549 tgctgtatgaattgac 565

RESULT 10
AAA26436
ID AAA26436 standard; cDNA; 1248 BP.
XX
AC AAA26436;
XX
DT 29-JUN-2000 (first entry)
XX
DE Human secreted protein gene 91 SEQ ID NO:101.
XX
KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW antiHIV; antiinflammatory; nootropic; neuroprotective; antiallergic;
KW osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
KW antipsoriatic; cardiant; gene therapy; cancer; neurological disorder;
KW immune disease; inflammation; blood disorder; tumour; ss.
XX
OS Homo sapiens.
XX
PN WO200006698-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-US17130.
XX
PR 30-JUL-1998; 98US-0094657.
PR 05-AUG-1998; 98US-0095486.
PR 06-AUG-1998; 98US-0095454.
PR 06-AUG-1998; 98US-0095455.
PR 12-AUG-1998; 98US-0096319.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;
XX
DR WPI; 2000-195282/17.
DR P-PSDB; AAY91541.
XX
XX New isolated human genes and the secreted polypeptides they encode,
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders -
XX
PS Claim 1; Page 435; 634pp; English.
XX
XX The polynucleotide sequences given in AAA26346 to AAA26458 encode the
CC human secreted proteins given in AAY91451 to AAY91691. The human secreted
CC proteins can have activities based on the tissues and cells they are
CC expressed in. Examples of the activities are: cytostatic;
CC immunosuppressive; antiHIV; antiinflammatory; nootropic; neuroprotective;
CC antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic;
CC antiasthma; antipsoriatic; and cardiant. The polynucleotides and their
CC corresponding secreted proteins are useful for preventing, treating or

CC ameliorating medical conditions, e.g. by protein or gene therapy. Also
CC pathological conditions can be diagnosed by determining the amount of the
CC proteins in a sample or by determining the presence of mutations in the
CC polynucleotides. Specific uses are described for each of the
CC polynucleotides, based on which tissues they are most highly expressed
CC in, and include developing products for the diagnosis or treatment of
CC cancer, tumours, neurodegenerative disorders, developmental abnormalities
CC and foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, inflammation,
CC allergies, Alzheimer's and behavioural disorders, schizophrenia,
CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
CC cardiovascular disorders, reproductive disorders, gastrointestinal
CC disorders, respiratory disorders and metabolic disorders. The proteins
CC or polynucleotides can also be used as food additives or preservatives.
CC The proteins are also useful for identifying their binding partners.
CC AAA26337 to AAA26345 and AA91450 are sequences used in the
CC exemplification of the present invention.

Sequence 1248 BP; 350 A; 412 C; 238 G; 248 T; 0 other;

Query Match	3.7%;	Score 17;	DB 21;	Length 1248;
Best Local Similarity	100.0%;	Pred. No. 28;		
Matches 17; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

OY .11 gagaatgggacagcatg 27
Db 328 gagaatgggacagcatg 344

RESULT 11

AAC76612 standard; cDNA; 1278 BP.

... AAC76612;

DT 08-FEB-2001 (first entry)

AA	DE	Human ORFX ORF2167 polynucleotide sequence	SEQ ID NO:4333.
AA	DE	Human ORFX ORF2167 polynucleotide sequence	SEQ ID NO:4333.

Human; open reading frame; ORF; detection; cytostatic; hepatotropic; vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; ss.

OS Homo sapiens.

PN W0200058473-A2.

PD 05-OCT-2000.

PF 31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 05-APR-1999; 99US-0127728.

XX

PA (CURA-) CURAGEN CORP.

AA Shimkets RA, Leach M;
PI

XX Pl

DR WPI; 2000-602362/57.
DR P-PSDB; AAB42403.

DR P-PSDB; AAB42403.

aa Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -

aa Claim 5; Page 3528-3529; 5507pp; English.
ps

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulninary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic; anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; cardiac; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, diabetes mellitus, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 1278 BP; 263 A; 386 C; 391 G; 238 T; 0 other;

Query Match	3.7%;	Score 17;	DB 21;	Length 1278;
Best Local Similarity	100.0%;	Pred. No. 28;		
Matches 17; Conservative	0;	Mismatches	0;	Indels 0;
				Gaps 0;

Qy	11	gagaatgggacagcatg	27
Db	1183	gaqaatgggacagcatg	1199

RESULT 12
AAC87123

AAC87123 standard; CDNA; 1648 BP.

AAC87123;

AA 20-APR-2001 (first entry)
DT

aa Nucleotide sequence of a plant auxin transporter polypeptide.
DE

xx Auxin; auxin transporter; AUX1; root gravitropism; transgenic plant;
kw plant phenotype; herbicide; ss.

aa Glycine max.
os

Key	Location/qualifiers
1	1.1
2	2.1
3	3.1
4	4.1
5	5.1
6	6.1
7	7.1
8	8.1
9	9.1
10	10.1
11	11.1
12	12.1
13	13.1
14	14.1
15	15.1
16	16.1
17	17.1
18	18.1
19	19.1
20	20.1
21	21.1
22	22.1
23	23.1
24	24.1
25	25.1
26	26.1
27	27.1
28	28.1
29	29.1
30	30.1
31	31.1
32	32.1
33	33.1
34	34.1
35	35.1
36	36.1
37	37.1
38	38.1
39	39.1
40	40.1
41	41.1
42	42.1
43	43.1
44	44.1
45	45.1
46	46.1
47	47.1
48	48.1
49	49.1
50	50.1
51	51.1
52	52.1
53	53.1
54	54.1
55	55.1
56	56.1
57	57.1
58	58.1
59	59.1
60	60.1
61	61.1
62	62.1
63	63.1
64	64.1
65	65.1
66	66.1
67	67.1
68	68.1
69	69.1
70	70.1
71	71.1
72	72.1
73	73.1
74	74.1
75	75.1
76	76.1
77	77.1
78	78.1
79	79.1
80	80.1
81	81.1
82	82.1
83	83.1
84	84.1
85	85.1
86	86.1
87	87.1
88	88.1
89	89.1
90	90.1
91	91.1
92	92.1
93	93.1
94	94.1
95	95.1
96	96.1
97	97.1
98	98.1
99	99.1
100	100.1

FT	CDS	2..1448
----	-----	---------

```

E1      /  cag
FT      /product= "auxin transporter"

```

AA
PN W0200078965-A2.

PD 28-DEC-2000.

21-JUN-2000; 2000WO-US17050.

22-JUN-1999; 99US-0140212.

AA
PA
(DUPO) DU PONT DE NEMOURS & CO E I.

XX 17

PI Cahoon RE, Weng Z;
XX
DR WPI: 2001-071396/08.
DR P-PSDB, AAB31241.
XX
PT New polynucleotide encoding protein that mediates effects of
PT phytohormone auxin, useful for screening for protein inhibitors for
PT potential as herbicides -
XX
PS Claim 2; Page 55; 69pp; English.
XX
CC The present sequence encodes an auxin transporter. The polypeptide
CC mediates he effects if the plant hormone auxin. The polypeptides of
CC the invention are homologous to the Arabidopsis auxin transporter AUX1,
CC which effects root gravitropism. The polynucleotides are used for
CC creating transgenic plants in which the auxin polypeptides are present
CC at higher or lower levels than normal plants. Plant phenotype can be
CC changed by specifically inhibiting expression of one or more genes by
CC antisense inhibition or cosuppression. The auxin polypeptides can be
CC used as a target to facilitate design and/or identification of inhibitors
CC of AUX1 that may be useful as herbicides. The auxin polynucleotides
CC may also be used as probes for gene mapping information that is useful
CC in plant breeding in order to develop lines with desired phenotypes.
XX
SQ Sequence 1648 BP; 404 A; 365 C; 364 G; 515 T; 0 other;

Query Match 3.7%; Score 17; DB 22; Length 1648;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaagagaaagaagt 17
|
Db 322 gaaagagaaagaagt 338

RESULT 13
AAC87122
ID AAC87122 standard; cDNA; 1775 BP.
XX
AC AAC87122;
XX
DT 20-APR-2001 (first entry)
XX
DE Nucleotide sequence of a plant auxin transporter polypeptide.
XX
KW Auxin; auxin transporter; AUX1; root gravitropism; transgenic plant;
KW plant phenotype; herbicide; ss.
XX
OS Glycine max.
XX
PN WO200078965-A2.
XX
PD 28-DEC-2000.
XX
PF 21-JUN-2000; 2000WO-US17050.
XX
PR 22-JUN-1999; 99US-0140212.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Cahoon RE, Weng Z;
XX
DR WPI: 2001-071396/08.
XX
PT New polynucleotide encoding protein that mediates effects of
PT phytohormone auxin, useful for screening for protein inhibitors for
PT potential as herbicides -
XX
PS Claim 2; Page 54-55; 69pp; English.
XX
CC The present sequence encodes an auxin transporter. The polypeptide
CC mediates he effects if the plant hormone auxin. The polypeptides of

CC the invention are homologous to the Arabidopsis auxin transporter AUX1,
CC which effects root gravitropism. The polynucleotides are used for
CC creating transgenic plants in which the auxin polypeptides are present
CC at higher or lower levels than normal plants. Plant phenotype can be
CC changed by specifically inhibiting expression of one or more genes by
CC antisense inhibition or cosuppression. The auxin polypeptides can be
CC used as a target to facilitate design and/or identification of inhibitors
CC of AUX1 that may be useful as herbicides. The auxin polynucleotides
CC may also be used as probes for gene mapping information that is useful
CC in plant breeding in order to develop lines with desired phenotypes.
XX
SQ Sequence 1775 BP; 447 A; 390 C; 376 G; 562 T; 0 other;

Query Match 3.7%; Score 17; DB 22; Length 1775;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaagagaaagaagt 17
|
Db 322 gaaagagaaagaagt 338

RESULT 14
AAC44199
ID AAC44199 standard; DNA; 2075 BP.
XX
AC AAC44199;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 41988.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.

PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.7%; Score 17; DB 21; Length 2075;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 cgatcgcttggcaaa 453
|||||
Db 1632 cgatcgcttggcaaa 1648

RESULT 15
AAQ79913
ID AAQ79913 standard; DNA; 3122 BP.
AC AAQ79913;

DT 01-AUG-1995 (first entry)

DE Lys-aminopeptidase pepN gene.

KW Lys-aminopeptidase; pepN gene; fermented food; cheese; ds.

OS Lactobacillus delbrueckii subsp. lactis.

FH Key Location/Qualifiers
FT CDS 316..2847
FT /*tag= a

PN EP633316-A.

PD 11-JAN-1995.

PF 30-JUN-1994; 94EP-0401497.

PR 01-JUL-1993; 93GB-0013586.

PA (EECE-) EEC EURO ECONOMIC COMMUNITY.

PI Klein JR, Plapp R;

DR WPI: 1995-038513/06.

DR P-PSDB; AAR67760.

PT Purified Lys-aminopeptidase pepN enzyme and pepN gene - useful
for prepn. of fermented foodstuff, esp. cheese.

PS Claim 8; Page 17-21; 41pp; English.

CC A new Lys-aminopeptidase, pepN, was isolated from L. delbrueckii
CC subsp. lactis WS87 (DSM 7290) and had the sequence given in
CC AAR67760. The pepN gene (AAQ79913) was isolated from a library
CC of DSM 7290 chromosomal DNA by screening for peptolytic activity
CC in Escherichia coli ER1562 transformants. The isolated gene
CC is used for recombinant pepN production.

SQ Sequence 3122 BP; 861 A; 745 C; 796 G; 720 T; 0 other;

Query Match 3.7%; Score 17; DB 16; Length 3122;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 cttctcaggacttc 196
|||||
Db 395 cttctcaggacttc 411

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2002, 00:18:18 ; Search time 81.38 Seconds
(without alignments)
1291.298 Million cell updates/sec

Title: US-09-621-781-1_COPY_1507_1970
Perfect score: 464
Sequence: 1 gaagagaagaagaatggga.....tttgcaaccgaaccgac 464

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	4.1	72928	3	US-09-009-913-1 Sequence 1, Appli
2	17	3.7	11298	1	US-07-869-933-31 Sequence 31, Appli
3	17	3.7	11298	1	US-08-201-879A-2 Sequence 2, Appli
4	17	3.7	11298	4	US-09-103-663-31 Sequence 31, Appli
5	16	3.4	1724	4	US-08-509-024-1 Sequence 1, Appli
6	16	3.4	1724	4	US-09-333-279-1 Sequence 1, Appli
7	16	3.4	1724	5	PCT-US96-12374-1 Sequence 1, Appli
8	16	3.4	1929	4	US-09-146-950-1 Sequence 28, Appli
9	16	3.4	2284	2	US-08-467-822-28 Sequence 28, Appli
10	16	3.4	2284	4	US-08-432-697-28 Sequence 28, Appli
11	16	3.4	2284	4	US-08-466-248-28 Sequence 2, Appli
12	16	3.4	2344	3	US-08-893-852A-2 Sequence 4, Appli
13	16	3.4	3177	1	US-08-042-747A-4 Sequence 7, Appli
14	16	3.4	3561	3	US-08-822-324-7 Sequence 4, Appli
15	16	3.4	5175	4	US-08-972-927-4 Sequence 5, Appli
16	16	3.4	10342	4	US-08-972-927-5 Sequence 1, Appli
17	16	3.4	15378	3	US-08-785-420-1 Sequence 20, Appli
18	16	3.4	246240	2	US-08-724-394A-20 Sequence 21, Appli
19	16	3.4	246240	2	US-08-724-394A-21 Sequence 22, Appli
20	16	3.4	246240	2	US-08-724-394A-22 Sequence 10, Appli
21	15	3.2	306	4	US-09-147-935A-10 Sequence 2, Appli
22	15	3.2	702	3	US-08-842-976-2 Sequence 2, Appli
23	15	3.2	702	3	US-09-213-397-2 Sequence 2, Appli
24	15	3.2	702	3	US-09-416-489-2 Sequence 47, Appli
25	15	3.2	822	4	US-08-896-164-47 Sequence 21, Appli
26	15	3.2	895	3	US-08-924-747-21 Sequence 21, Appli
27	15	3.2	895	4	US-09-247-373B-21 Sequence 21, Appli

28	15	3.2	895	4	US-09-296-715-21 Sequence 21, Appli
29	15	3.2	1058	3	US-09-156-807-1 Sequence 15, Appli
30	15	3.2	1189	3	US-08-961-083-15 Sequence 9, Appli
31	15	3.2	1569	3	US-08-821-984-9 Sequence 4, Appli
32	15	3.2	1569	4	US-09-329-749-9 Sequence 104, App
33	15	3.2	1659	1	US-08-548-509-4 Sequence 104, App
34	15	3.2	1697	1	US-08-181-271A-104 Sequence 104, App
35	15	3.2	1697	1	US-08-449-315-104 Sequence 104, App
36	15	3.2	1697	1	US-08-444-803-104 Sequence 104, App
37	15	3.2	1697	1	US-08-449-043-104 Sequence 104, App
38	15	3.2	1697	1	US-08-456-265A-104 Sequence 104, App
39	15	3.2	1697	1	US-08-455-416-104 Sequence 104, App
40	15	3.2	1697	1	US-08-454-876-104 Sequence 104, App
41	15	3.2	1697	2	US-08-457-364-104 Sequence 104, App
42	15	3.2	1697	2	US-08-456-262-104 Sequence 104, App
43	15	3.2	1697	2	US-08-456-240-104 Sequence 104, App
44	15	3.2	1697	2	US-08-455-736-104 Sequence 104, App
45	15	3.2	1697	2	US-08-455-736-104 Sequence 104, App

ALIGNMENTS

RESULT 1
US-09-009-913-1
; Sequence 1, Application US/09009913
; Patent No. 6087485
; GENERAL INFORMATION:
; APPLICANT: Axy's Pharmaceuticals, Inc.
; TITLE OF INVENTION: Asthma Related Genes
; NUMBER OF SEQUENCES: 339
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,913
; FILING DATE: 21-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: SEQ-4P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3231
; TELEFAX: 650-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72928 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-09-009-913-1

Query Match 4.1%; Score 19; DB 3; Length 72928;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 aaagagaagaagaatggga 20

Db 46222 AAAGAAAGAGATGGGA 46240

RESULT 2

US-07-869-933-31
; Sequence 31, Application US/07869933
; Patent No. 5770396

GENERAL INFORMATION:

APPLICANT: KINET, Jean-Pierre
TITLE OF INVENTION: ISOLATION, CHARACTERIZATION, AND USE OF
TITLE OF INVENTION: THE HUMAN B SUBUNIT OF THE HIGH AFFINITY RECEPTOR FOR
TITLE OF INVENTION: IMMUNOGLOBULIN
NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA

ZIP: 22313-0299

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/869,933
FILING DATE: 19920416
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 40399/154 NIHD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109

TELEX: 899149

INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:
LENGTH: 11298 base pairs
TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: linear

ORIGINAL SOURCE:
ORGANISM: homo sapien

STRAIN: FCRI beta

US-07-869-933-31

Query Match

Best Local Similarity 3.7%; Score 17; DB 1; Length 11298;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 agagaaagagaatggga 20

Db 9161 AGAGAAAGAGATGGGA 9177

RESULT 3

US-08-201-879A-2
; Sequence 2, Application US/08201879A
; Patent No. 5807988

GENERAL INFORMATION:

APPLICANT: KINET, Jean-Pierre
APPLICANT: JOUVIN, Marie-Helene
TITLE OF INVENTION: ISOLATION, CHARACTERIZATION, AND USE OF
TITLE OF INVENTION: THE HUMAN B SUBUNIT OF THE HIGH AFFINITY RECEPTOR FOR
TITLE OF INVENTION: IMMUNOGLOBULIN E
NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/201,879A
FILING DATE: 24-FEB-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/869,933
FILING DATE: 16-APR-1992
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/03419
FILING DATE: 16-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 40399/234/NIHD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 11298 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

STRAIN: FCRI beta

FEATURE:

NAME/KEY: CDS

LOCATION: join(456..511, 1381..1510, 2026..2160, 4475..4531,
LOCATION: 5079..5237, 5640..5738, 7224..7319)

US-08-201-879A-2

QY 4 agagaaagagaatggga 20

Db 9161 AGAGAAAGAGATGGGA 9177

RESULT 4

US-09-103-663-31
; Sequence 31, Application US/09103663D
; Patent No. 6171803

GENERAL INFORMATION:

APPLICANT: Kinet et al.
TITLE OF INVENTION: Isolation, characterization, and use of the human beta
TITLE OF INVENTION: subunit of the high affinity receptor for
TITLE OF INVENTION: immunoglobulin E.
FILE REFERENCE: 50490

CURRENT APPLICATION NUMBER: US/09/103,663D
CURRENT FILING DATE: 1998-06-23

EARLIER APPLICATION NUMBER: 07/869,933
EARLIER FILING DATE: 1992-04-16

NUMBER OF SEQ ID NOS: 35
SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 31

LENGTH: 11298
TYPE: DNA

ORGANISM: Homo sapiens
US-09-103-663-31

Query Match 3.7%; Score 17; DB 4; Length 11298;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 agagaaagaagaatggga 20
|||||
Db 9161 agagaaagaagaatggga 9177

RESULT 5
US-08-509-024-1/c
; Sequence 1, Application US/08509024B
; Patent No. 6291207
; GENERAL INFORMATION:
; APPLICANT: SPEAR, Patricia G.
; APPLICANT: MONTGOMERY, Rebecca I.
; TITLE OF INVENTION: HERPES VIRUS ENTRY RECEPTOR PROTEIN
; FILE REFERENCE: 0290-1
; CURRENT APPLICATION NUMBER: US/08/509,024B
; CURRENT FILING DATE: 1995-07-25
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1724
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-509-024-1

Query Match 3.4%; Score 16; DB 4; Length 1724;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaagagaagaagaat 16
|||||
Db 198 GAAAGAGAAAGAGAAT 183

RESULT 6
US-09-333-279-1/c
; Sequence 1, Application US/09333279
; Patent No. 6303336
; GENERAL INFORMATION:
; APPLICANT: SPEAR, Patricia G.
; APPLICANT: MONTGOMERY, Rebecca I.
; TITLE OF INVENTION: HERPES VIRUS ENTRY RECEPTOR PROTEIN
; FILE REFERENCE: 0290-1
; CURRENT APPLICATION NUMBER: US/09/333,279
; CURRENT FILING DATE: 1999-06-15
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1724
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-333-279-1

Query Match 3.4%; Score 16; DB 4; Length 1724;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaagagaagaagaat 16
|||||
Db 198 GAAAGAGAAAGAGAAT 183

RESULT 7
PCT-US96-12374-1/c
; Sequence 1, Application PC/TUS9612374
; GENERAL INFORMATION:

; APPLICANT: Northwestern University
; TITLE OF INVENTION: Herpes Virus Entry Mediator
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dressler, Goldsmith, Milnamow & Katz, Ltd.
; STREET: 180 N. Stetson, Suite 4700
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/12374
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Northrup, Thomas E.
; REGISTRATION NUMBER: 33,268
; REFERENCE/DOCKET NUMBER: NOR3446P020PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5400
; TELEFAX: (312) 616-5460
; TELEX: --
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1724 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 294..1145
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 294..1142
PCT-US96-12374-1

Query Match 3.4%; Score 16; DB 5; Length 1724;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaagagaagaagaat 16
|||||
Db 198 GAAAGAGAAAGAGAAT 183

RESULT 8
US-09-146-950-1/c
; Sequence 1, Application US/09146950A
; Patent No. 6287808
; GENERAL INFORMATION:
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRY-MEDIATOR-RELATED
; FILE REFERENCE: 09404/057001
; CURRENT APPLICATION NUMBER: US/09/146,950A
; CURRENT FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1929
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (297)...(875)

US-09-146-950-1

Query Match 3.4%; Score 16; DB 4; Length 1929;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaagagaagaagaat 16
|||||
Db 195 GAAAGAGAAAGAGAAT 180

RESULT 9

US-08-467-822-28
; Sequence 28, Application US/08467822
; Patent No. 5843460

; GENERAL INFORMATION:

; APPLICANT: Labigne, Agnes

; APPLICANT: Sauerbaum, Sebastien

; APPLICANT: Ferrero, Richard L.

; APPLICANT: Thiberge, Jean-Michel

; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST

; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE

; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID

; TITLE OF INVENTION: POLYPEPTIDES

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

; ADDRESSEE: Dunner

; STREET: 1300 I Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/467,822

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/447,177

; FILING DATE: 19-MAY-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/432,697

; FILING DATE: 02-MAY-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Meyers, Kenneth J.

; REGISTRATION NUMBER: 25,146

; REFERENCE/DOCKET NUMBER: 03495.0137-02000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 408-4000

; TELEFAX: (202) 408-4400

; INFORMATION FOR SEQ ID NO: 28:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2284 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; US-08-467-822-28

Db 300 GATCGCTTTGGCAA 315
|||||

RESULT 10

US-08-432-697-28
; Sequence 28, Application US/08432697
; Patent No. 6248330

; GENERAL INFORMATION:

; APPLICANT: Labigne, Agnes

; APPLICANT: Sauerbaum, Sebastien

; APPLICANT: Ferrero, Richard L.

; APPLICANT: Thiberge, Jean-Michel

; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST

; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE

; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID

; TITLE OF INVENTION: POLYPEPTIDES

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

; ADDRESSEE: Dunner

; STREET: 1300 I Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/432,697

; FILING DATE: 02-MAY-1995

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Meyers, Kenneth J.

; REGISTRATION NUMBER: 25,146

; REFERENCE/DOCKET NUMBER: 03495.0137-00000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 408-4000

; TELEFAX: (202) 408-4400

; INFORMATION FOR SEQ ID NO: 28:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2284 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; US-08-432-697-28

Query Match 3.4%; Score 16; DB 4; Length 2284;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 gatcgcttttgcaaa 453
|||||
Db 300 GATCGCTTTTGCAAA 315

RESULT 11

US-08-466-248-28
; Sequence 28, Application US/08466248
; Patent No. 6258359

; GENERAL INFORMATION:

; APPLICANT: Labigne, Agnes

; APPLICANT: Sauerbaum, Sebastien

; APPLICANT: Ferrero, Richard L.

; APPLICANT: Thiberge, Jean-Michel

; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST

; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE

; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID

Query Match 3.4%; Score 16; DB 2; Length 2284;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 gatcgcttttgcaaa 453

```

; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,248
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2284 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-466-248-28

Query Match          3.4%; Score 16; DB 4; Length 2284;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 gatcgctttggcaaa 453
      |||||||||||||||
Db 300 GATCGCTTTGGCAAA 315

RESULT 12
US-08-893-852A-2
; Sequence 2, Application US/08893852A
; Patent No. 6080558
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

```

; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,852A
; FILING DATE: Filed Herewith
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0341 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2344 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TMLR3DT01
; CLONE: 508302
; US-08-893-852A-2

Query Match          3.4%; Score 16; DB 3; Length 2344;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 402 atgaacgcgtgcct 417
      |||||||||||||||
Db 119 ATGACCGCGCTGGCCT 134

RESULT 13
US-08-042-747A-4/c
; Sequence 4, Application US/08042747A
; Patent No. 5487969
; GENERAL INFORMATION:
; APPLICANT: Eberle, Richard
; APPLICANT: Black, Darla
; APPLICANT: Scinicariello, Franco
; APPLICANT: Hilliard, Julia K.
; TITLE OF INVENTION: Cloning and Amplification of Monkey B
; TITLE OF INVENTION: Virus Genes
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cox & Smith Incorporated
; STREET: 112 East Pecan Street, Suite 2000
; CITY: San Antonio
; STATE: Texas
; COUNTRY: USA
; ZIP: 78205
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/042,747A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haymond, W. Bradley
; REGISTRATION NUMBER: 35186
; REFERENCE/DOCKET NUMBER: S-0072.179
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 210-554-5500
; TELEFAX: 210-226-8395
; TELEX: 767609
```


; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3177 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 269..2941
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..249
; US-08-042-747A-4

Query Match 3.4%; Score 16; DB 1; Length 3177;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 368 catcacgtcccgagc 383
|||||
Db 1933 CATCACGTCCCGAGC 1918

RESULT 14
US-08-822-324-7/C
; Sequence 7, Application US/08822324
; Patent No. 6129917
; GENERAL INFORMATION:
; APPLICANT: Potempa, Jan S.
; APPLICANT: Travis, James
; APPLICANT: Genco, Caroline A.
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS COMPRISING
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,324
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 536
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/013,945
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 103-95 WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 488-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3561 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS

; LOCATION: 1336..2862
; US-08-822-324-7

Query Match 3.4%; Score 16; DB 3; Length 3561;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 tctcaccttctcag 188
|||||
Db 826 TCTCACCTTCTCAG 811

RESULT 15
US-08-972-927-4
; Sequence 4, Application US/08972927
; Patent No. 616290
; GENERAL INFORMATION:
; APPLICANT: Rea, Phillip A
; APPLICANT: Lu, Yu-Ping
; APPLICANT: Li, Ze-Sheng
; TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN
; TITLE OF INVENTION: PLANTS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: One Commerce Square, 2005 Market Street, 22nd
; STREET: Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: US
; ZIP: 19103-7086
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/972,927
; FILING DATE: 18-NOV-1997
; CLASSIFICATION: 800
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,040
; FILING DATE: 18-NOV-1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/061,328
; FILING DATE: 08-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Doyle leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9596-1202
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-965-1284
; TELEFAX: 215-567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5175 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-972-927-4

Query Match 3.4%; Score 16; DB 4; Length 5175;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 149 gagaatttcagtgtg 164
|||||
Db 4239 GAGAAITTCAGTGTG 4254

'Sat Apr 6 11:03:27 2002

us-09-621-781-1_copy_1507_1970.rni

Page 7

Search completed: April 6, 2002, 06:18:07
Job time: 21589 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2002, 00:07:08 ; Search time 1415.1 Seconds
(without alignments)
3523.455 Million cell updates/sec

Title: US-09-621-781-1_COPY_1507_1970

Perfect score: 464

Sequence: 1 gaaagagaaagagaatggga.....tttgcaaacgcgaaccgcgac 464

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_esttpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	35	7.5	461	10	AA920479	AA920479 vy53c08.r
2	35	7.5	542	10	AW762360	AW762360 ur56h03.y
3	35	7.5	574	11	BG100299	BG100299 ux84f03.y
4	35	7.5	575	10	AA472295	AA472295 vh01c05.r
5	35	7.5	596	10	BE573255	BE573255 601333152
6	35	7.5	615	11	W62865	W62865 md85f08.r1
7	35	7.5	623	10	AM320574	AM320574 uo20f09.y
8	35	7.5	708	10	BE573350	BE573350 601333015
9	35	7.5	908	11	BF182426	BF182426 601804187
10	35	7.5	1136	10	BE573733	BE573733 601333482
11	26	5.6	446	10	AA684447	AA684447 vm64b07.s
12	26	5.6	573	10	AA104251	AA104251 mo50b03.r

13	21	4.5	1077	10	BE614567	BE614567 601503841
14	20	4.3	454	10	AM634278	AM634278 b118901.w
15	20	4.3	469	10	AM634288	AM634288 b118h01.w
16	20	4.3	471	10	AA789164	AA789164 ag58b01.s
17	20	4.3	492	13	AQ536955	AQ536955 RPCT-11-3
18	20	4.3	515	10	A1759543	A1759543 ETESTea29
19	20	4.3	594	13	AQ537404	AQ537404 RPCT-11-3
20	20	4.3	936	11	B1113465	B1113465 602899696
21	19	4.1	191	10	AV069346	AV069346 AV069346
22	19	4.1	230	10	BB017230	BB017230 BB017230
23	19	4.1	249	10	BB252955	BB252955 BB252955
24	19	4.1	330	10	AV212169	AV212169 AV212169
25	19	4.1	400	11	BF715432	BF715432 mab07g04.
26	19	4.1	432	13	AZ817262	AZ817262 2M0086J03
27	19	4.1	540	11	BG363063	BG363063 sac09g06.
28	19	4.1	620	13	AZ937634	AZ937634 2M0195F22
29	19	4.1	632	13	AZ949946	AZ949946 2M0213H17
30	19	4.1	746	10	BE038724	BE038724 AB04F03.A
31	19	4.1	836	10	BE782858	BE782858 601472368
32	18	3.9	79	10	A1966441	A1966441 SC39b06.y
33	18	3.9	150	13	BH055024	BH055024 RPCT-24-2
34	18	3.9	173	13	AQ389139	AQ389139 RPCT11-15
35	18	3.9	228	11	BF716107	BF716107 saal14f03.
36	18	3.9	229	10	A1506763	A1506763 vm58c08.x
37	18	3.9	248	10	AA596509	AA596509 vm58c08.r
38	18	3.9	253	13	AZ621106	AZ621106 1M0454D08
39	18	3.9	262	10	BE058354	BE058354 sn14h11.y
40	18	3.9	263	11	BG049447	BG049447 CVL_20_A1
41	18	3.9	280	10	AW134889	AW134889 UI-H-B11-
42	18	3.9	363	10	BE363395	BE363395 WSL_62_D1
43	18	3.9	364	13	AZ553147	AZ553147 RPCT-23-2
44	18	3.9	374	13	AZ655779	AZ655779 1M0530L20
45	18	3.9	378	10	A1494185	A1494185 t114g10.y

ALIGNMENTS

RESULT 1
AA920479 461 bp mRNA EST 20-APR-1998
LOCUS vy53c08.r1 Stratagene mouse lung 937302 Mus musculus cDNA clone
DEFINITION IMAGE:1299182 5' similar to gb:X51829 Mouse myeloid differentiation
primary response mRNA encoding (MOUSE);, mRNA sequence.

ACCESSION
AA920479
AA920479.1 GI:3067258

VERSION
KEYWORDS
SOURCE
ORGANISM

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

The WashU-HHMI Mouse EST Project

TITLE
JOURNAL
COMMENT

Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:680230

Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 348.
Location/Qualifiers

FEATURES
source 1..461


```
/organism="Mus musculus"
/strain="C57BL/6 x CBA"
/db_xref="taxon:10090"
/clone="IMAGE:1299182"
/clone_lib="Stratagene mouse lung 937302"
/sex="female"
/tissue_type="lung"
/dev_stage="6-8 month old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dT. 6-8 month old female lung and 1.5 year old male lung
were source of mRNA. Average insert size: 1.5 kb; Uni-ZAP
XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'
adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

BASE COUNT      102 a      157 c      125 g      77 t

ORIGIN

Query Match
Best Local Similarity  7.5%; Score 35; DB 10; Length 461;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  373  cgtcccgagcccgagcctgagggcgacatgaac 407
|||||
Db    66  CGTCCCGAGCCCGACGCTGAGGGCGACATGAAC 100

RESULT  2
AW762360      542 bp      mRNA      EST      04-MAY-2000
LOCUS      ur56h03.y1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:3154325 5'
DEFINITION similar to gb:X51829 Mouse myeloid differentiation primary response
            mRNA encoding (MOUSE);, mRNA sequence.
ACCESSION  AW762360
VERSION    AW762360.1 GI:7694285
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus

REFERENCE
AUTHORS     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE       1 (bases 1 to 542)
JOURNAL     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Other_ESTs: ur56h03.x1
            Contact: Robert Strausberg, Ph.D.
            Email: cgabs-remail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
            CDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml

MG1:1057081
Seq primer: -40RP from gibco
High quality sequence stop: 374.
Location/Qualifiers
1..542
/organism="Mus musculus"
/strain="129 - C57/B6 - FVBN"
/db_xref="taxon:10090"
/clone="IMAGE:3154325"
/clone_lib="NCI_CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
```

```
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22, 37-43 (1999)."

BASE COUNT      114 a      171 c      159 g      98 t

ORIGIN

Query Match
Best Local Similarity  7.5%; Score 35; DB 10; Length 542;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  373  cgtcccgagcccgagcctgagggcgacatgaac 407
|||||
Db    87  CGTCCCGAGCCCGACGCTGAGGGCGACATGAAC 121

RESULT  3
BG100299      574 bp      mRNA      EST      29-JAN-2001
LOCUS      ux84f03.y1 McCarrey Eddy type B spermatogonia Mus musculus cDNA
DEFINITION clone IMAGE:3655420 5' similar to SW:MY16 MOUSE P17564 MYELOID
            DIFFERENTIATION PRIMARY RESPONSE PROTEIN MYD16. ;, mRNA sequence.
ACCESSION  BG100299
VERSION    BG100299.1 GI:12595616
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus

REFERENCE
AUTHORS     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE       1 (bases 1 to 574)
JOURNAL     Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
            The Washu-NCI Mouse EST Project 1999
            Unpublished (1999)
            Contact: Marra M/Washu-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MG1:1417724
            Seq primer: Primer name ambiguous
            High quality sequence stop: 392.
            Location/Qualifiers
            1..574
            /organism="Mus musculus"
            /strain="CD-1"
            /db_xref="taxon:10090"
            /clone="IMAGE:3655420"
            /clone_lib="McCarrey Eddy type B spermatogonia"
            /sex="male"
            /tissue_type="type B spermatogonia, pooled from multiple
            mice"
            /dev_stage="8 day"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: testis; Vector: pBluescript SK+ (Stratagene
            ); Site_1: XhoI; Site_2: EcoRI; cDNA oligo dT-primed
            [5'-(GA)10-ACTAGTCCTGAGTTTCTTTT-3'] and directionally
            cloned using 5' linkers 5'-AATTGGCAGAG-3' and
            5'-CTCGTCCG-3'. Size selection of >400bp material gives
            average insert size ranging from 1-2 kb. Library was mass
            excised (from lambda-UniZAP-XR) and resulting
            single-stranded phagemids were prepped and transformed
            into DH10B. Library contains 968 recombinants.
            References: J. Androl. 20:635-639 and Gene 25:263-269.
            Library constructed and donated by J. McCarrey, Ph.D.
            (Southwest Foundation for Biomedical Research, Dept. of
            Genetics); excision done by E.M. Eddy, Ph.D. (National
```

Institutes of Health, National Institute of Environmental Health Sciences). Original lambda-based library is available through ATCC, catalog #63417."

BASE COUNT 133 a 179 c 159 g 100 t 3 others
ORIGIN

Query Match 7.5%; Score 35; DB 11; Length 574;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 373 cgtcccgagcccccagcctgagggcgacatgaac 407
|||||
Db 51 CGTCCCGAGCCCCACGCTGAGGGCGACATGAAC 85

RESULT 4
AA472295 575 bp mRNA EST 18-JUN-1997
LOCUS
DEFINITION Vh01C05.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
IMAGE:874184 5' similar to gb:X51829 Mouse myeloid differentiation
primary response mRNA encoding (MOUSE);, mRNA sequence.

ACCESSION AA472295 GI:2200286
VERSION
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 575)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:513664

Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 377.

FEATURES
source Location/Qualifiers

1..575
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:874184"
/clone_lib="Soares_mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
RI: 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5'
TGTACCAATCTGAGTGGAGCGCGCGAATGTTT

T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru KO, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Patima
Bonaldo."

BASE COUNT 127 a 183 c 161 g 104 t
ORIGIN

Query Match 7.5%; Score 35; DB 10; Length 575;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 373 cgtcccgagcccccagcctgagggcgacatgaac 407
|||||
Db 93 CGTCCCGAGCCCCACGCTGAGGGCGACATGAAC 127

RESULT 5
BE573255 596 bp mRNA EST 15-AUG-2000
LOCUS
DEFINITION 601333152F2 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3710629 5',
mRNA sequence.

ACCESSION BE573255 GI:9816975
VERSION
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 596)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov

Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM8933 row: g column: 14

High quality sequence stop: 542.

FEATURES
source Location/Qualifiers

1..596
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3710629"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 132 a 199 c 163 g 102 t
ORIGIN

Query Match 7.5%; Score 35; DB 10; Length 596;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 373 cgtcccgagcccccagcctgagggcgacatgaac 407
|||||
Db 82 CGTCCCGAGCCCCACGCTGAGGGCGACATGAAC 116

RESULT 6
W62865 615 bp mRNA EST 07-JUN-1996
LOCUS
DEFINITION md85f08.r1 Soares embryo NbME13.5 14.5 Mus musculus cDNA
clone IMAGE:375207 5' similar to gb:X51829 Mouse myeloid
differentiation primary response mRNA encoding (MOUSE);, mRNA
sequence.

ACCESSION	W62865
VERSION	W62865.1
KEYWORDS	GI:1369614
SOURCE	EST.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 615) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MG1:236639 Seq primer: ETPrimer High quality sequence stop: 344.
FEATURES	Location/Qualifiers 1..615 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:375207" /clone_lib="Soares mouse embryo NbME13.5 14.5" /sex="unknown" /tissue_type="embryo" /dev_stage="13.5-14.5dpc total fetus" /lab_host="DH10B" /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5', TGTTACCAATCTGAAGTGGAGCGCCGGAATTGTGTGTGTGTGTGTGT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT	145 a 190 c 165 g 115 t.
ORIGIN	
Query Match	7.5%; Score 35; DB 11; Length 615;
Best Local Similarity	100.0%; Pred. No. 3.2e-07;
Matches	35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY	373 cgtccccgagcccaagcctgaggcgacatgaac 407
Dp	92 CGTCCCCGAGCCCCACGCCTGAGGGCGACATGAAC 126
RESULT	7
LOCUS	AM320574
DEFINITION	AM320574 623 bp mRNA EST 25-JAN-2000
ACCESSION	U020F09.y1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:2609705 5'
VERSION	similar to gb:X51829 Mouse myeloid differentiation primary response
KEYWORDS	mRNA encoding (MOUSE); mRNA sequence.
SOURCE	AM320574.1 GI:6750118 EST. house mouse.

ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE	1 (bases 1 to 623)
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Jeffrey Green M.D. CDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
FEATURES	MG1:1023565 Seq primer: -40RP from Gibco High quality sequence stop: 434. Location/Qualifiers 1..623 /organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090" /clone="IMAGE:2609705" /clone_lib="NCI-CGAP_Mam6" /sex="female, virgin" /tissue_type="infiltrating ductal carcinoma" /dev_stage="5 months" /lab_host="DH10B" /note="Organ: mammary; Vector: PCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT	137 a 197 c 173 g 115 t 1 others
ORIGIN	
Query Match	7.5%; Score 35; DB 10; Length 623;
Best Local Similarity	100.0%; Pred. No. 3.2e-07;
Matches 35; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY 373	cgtcccgagccccacgcctgaggcgacatgaac 407
Db 119	CGTCCCGAGCCCCACGCCTGAGGGCAGACATGAAC 153
RESULT 8	
BE573350	708 bp mRNA EST 15-AUG-2000
LOCUS	601333015F2 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3710762 5', mRNA sequence.
DEFINITION	BE573350
ACCESSION	BE573350.1 GI:9817070
VERSION	EST.
KEYWORDS	house mouse.
SOURCE	Mus musculus
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 708) NIH-MGC http://mgc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Jeffrey Green M.D. CDNA Library Preparation: Life Technologies, Inc. CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

http://image.llnl.gov
Plate: LAM8933 row: m column: 03
High quality sequence stop: 649.
Location/Qualifiers

FEATURES
source

1. 708
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3710762"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT 165 a 213 c 208 g 122 t
ORIGIN

Query Match 7.5%; Score 35; DB 10; Length 708;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 373 cgtccccgagcccgctgagggcgacatgaac 407
|||||
Db 73 CGTCCCGAGCCCCACGCTGAGGGCGACATGAAC 107

RESULT 9
BF182426 908 bp mRNA EST 31-OCT-2000
LOCUS 601804187F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4035261 5',
DEFINITION mRNA sequence.
ACCESSION BF182426
VERSION BF182426.1 GI:11060569
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 908)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM9308 row: m column: 22
High quality sequence stop: 646.
Location/Qualifiers

REFERENCE 1. 908
AUTHORS /organism="Mus musculus"
TITLE /strain="C57/B6"
JOURNAL /db_xref="taxon:10090"
COMMENT /clone="IMAGE:4035261"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

FEATURES
source

1. 908
/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone="IMAGE:4035261"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

BASE COUNT 220 a 271 c 264 g 153 t
ORIGIN

Query Match 7.5%; Score 35; DB 11; Length 908;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 373 cgtccccgagcccgctgagggcgacatgaac 407
|||||
Db 92 CGTCCCGAGCCCCACGCTGAGGGCGACATGAAC 126

RESULT 10
BE573733 1136 bp mRNA EST 15-AUG-2000
LOCUS 601333482F2 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3710924 5',
DEFINITION mRNA sequence.
ACCESSION BE573733
VERSION BE573733.1 GI:9817453
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1136)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM8934 row: c column: 21
High quality sequence stop: 635.
Location/Qualifiers

REFERENCE 1. 1136
AUTHORS /organism="Mus musculus"
TITLE /strain="FVB/N"
JOURNAL /db_xref="taxon:10090"
COMMENT /clone="IMAGE:3710924"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

FEATURES
source

1. 1136
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3710924"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT 300 a 363 c 311 g 161 t
ORIGIN

Query Match 7.5%; Score 35; DB 10; Length 1136;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 373 cgtccccgagcccgctgagggcgacatgaac 407
|||||
Db 81 CGTCCCGAGCCCCACGCTGAGGGCGACATGAAC 115

RESULT 11
AA684447 446 bp mRNA EST 09-DEC-1997
LOCUS vm64b07.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
DEFINITION IMAGE:1002997 5' similar to gb:X51829 Mouse myeloid differentiation

ACCESSION AA684447 primary response mRNA encoding (MOUSE);, mRNA sequence.
VERSION AA684447.1 GI:2671033
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 446)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The Washu-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
Washu-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:567213
High quality sequence stop: 280.
Location/Qualifiers
1..446
/organism="Mus musculus"
/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/clone="IMAGE:1002997"
/clone_lib="Knowles Solter mouse 2 cell"
/tissue_type="embryo"
/dev_stage="2-cell"
/lab_host="DH10B"
/note="Organ: embryo; Vector: pBluescribe (modified); site_1: MluI; site_2: SalI; Cloned unidirectionally from mRNA prepared from 13,500 2-cell stage embryos. Primer: SalI(dT): 5'-CGGTCGACCGTCGACCGTTTCTTTT-3', CDNAS were cloned into the MluI/SalI sites of a modified pBluescribe vector using commercial linkers (NMB). Average insert size: 1.2 kb."
BASE COUNT 79 a 123 c 154 g 90 t
ORIGIN
Query Match 5.6%; Score 26; DB 10; Length 446;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 427 tccggaccacgacgcttggcaa 452
|||||
Db 294 TCCGACCCACGATCGCTTTGGCAA 269
RESULT 12
AA104251 573 bp mRNA EST 29-OCT-1996
LOCUS AA104251
DEFINITION mo50D03.r1 Life Tech mouse embryo 10 5dpc 10665016 Mus musculus
CDNA clone IMAGE:556973 5' similar to gb:X51829 Mouse myeloid differentiation primary response mRNA encoding (MOUSE);, mRNA
sequence.
ACCESSION AA104251
VERSION AA104251.1 GI:1650409
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 573)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The Washu-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
Washu-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:337765
Seq primer: -28M13 rev1 from Amersham
High quality sequence stop: 246.
Location/Qualifiers
1..573
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:556973"
/clone_lib="Life Tech mouse embryo 10 5dpc 10665016"
/tissue_type="embryo"
/dev_stage="10.5dpc embryos"
/lab_host="DH10B"
/note="Organ: whole embryo; Vector: PCMV-SPORT2; Site_1: SalI; site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. 10.5dpc embryos. PCMV-SPORT2 vector."
BASE COUNT 123 a 178 c 156 g 115 t 1 others
ORIGIN
Query Match 5.6%; Score 26; DB 10; Length 573;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 427 tccggaccacgacgcttggcaa 452
|||||
Db 147 TCCGACCCACGATCGCTTTGGCAA 172
RESULT 13
BE614567 1077 bp mRNA EST 20-OCT-2000
LOCUS BE614567
DEFINITION 601503841T1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3905554 3', mRNA sequence.
ACCESSION BE614567
VERSION BE614567.1 GI:9896164
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 1077)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LHAM9713 row: a column: 11
High quality sequence start: 60
High quality sequence stop: 682.
Location/Qualifiers
FEATURES

```

source
1. .1077
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3905554"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: uterus; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT      289 a      268 g      235 t
ORIGIN

Query Match      4.5%; Score 21; DB 10; Length 1077;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      20 acagcatgtgactgcctgatg 40
      |||||||||||||||||||
Db      761 ACAGCATGTGACTGCCTGATG 781

RESULT 14
AM634278      454 bp      mRNA      EST      26-APR-2001
LOCUS      b118q01.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
DEFINITION      laevis cDNA clone PBX0018G01 5', mRNA sequence.
ACCESSION      AM634278
VERSION      AM634278.1 GI:7391359
KEYWORDS      EST.
SOURCE      African clawed frog.
ORGANISM      Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 454)
Blackshear,P.J., Lai,W.S., Thorn,J.M., Kennington,E.A., Staffa,N.G.
Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman
,J.W., Bonaldo,M.F. and Soares,M.B.
The NIEHS Xenopus maternal EST project: interim analysis of the
first 13,879 ESTs from unfertilized eggs
Gene 267 (1), 71-87 (2001)
21211403
TITLE      Contact: Perry J. Blackshear
JOURNAL      Office of Clinical Research and Laboratory of Signal Transduction
MEDLINE      National Institute of Environmental Health Sciences
COMMENT      A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
USA
Tel: 919 541-4899
Fax: 919 541-4571
Email: black009@niehs.nih.gov
Clone is available through Research Genetics, Inc., 2130 Memorial
Parkway, Huntsville, AL 35901
phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email
cdna@resgen.com
DNA Sequencing and analyses performed by National Institutes of
Health Intramural Sequencing Center (NISC).
PCR Primers
FORWARD: TGTAAACGACGGCCAGT
BACKWARD: CAGGAACAGCTATGACC
Plate: 0018 row: G column: 01
Seq primer: T7 primer.
location/Qualifiers
1. .454
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="PBX0018G01"
/clone_lib="Blackshear/Soares normalized Xenopus egg
library"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
```

```

/dev_stage="unfertilized egg"
/lab_host="DH10B"
/notes="Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery', Genome Research
6:791-806, 1996. The first strand synthesis used a
NotI-dT18 primer; double stranded cDNAs were ligated to
EcoRI adapters, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested pT7T3-Pac vector.
The library contained approximately 7.2 x 105
recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT      130 a      69 c      134 g      121 t
ORIGIN

Query Match      4.3%; Score 20; DB 10; Length 454;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      4 agagaagaagatggagacag 23
      |||||||||||||||||||
Db      78 AGAGAAAGAGAAATGGGACAG 97

RESULT 15
AM634288      469 bp      mRNA      EST      26-APR-2001
LOCUS      b118h01.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
DEFINITION      laevis cDNA clone PBX0018H01 5', mRNA sequence.
ACCESSION      AM634288
VERSION      AM634288.1 GI:7391369
KEYWORDS      EST.
SOURCE      African clawed frog.
ORGANISM      Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 469)
Blackshear,P.J., Lai,W.S., Thorn,J.M., Kennington,E.A., Staffa,N.G.
Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman
,J.W., Bonaldo,M.F. and Soares,M.B.
The NIEHS Xenopus maternal EST project: interim analysis of the
first 13,879 ESTs from unfertilized eggs
Gene 267 (1), 71-87 (2001)
21211403
TITLE      Contact: Perry J. Blackshear
JOURNAL      Office of Clinical Research and Laboratory of Signal Transduction
MEDLINE      National Institute of Environmental Health Sciences
COMMENT      A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
USA
Tel: 919 541-4899
Fax: 919 541-4571
Email: black009@niehs.nih.gov
Clone is available through Research Genetics, Inc., 2130 Memorial
Parkway, Huntsville, AL 35901
phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email
cdna@resgen.com
DNA Sequencing and analyses performed by National Institutes of
Health Intramural Sequencing Center (NISC).
PCR Primers
FORWARD: TGTAAACGACGGCCAGT
BACKWARD: CAGGAACAGCTATGACC
Plate: 0018 row: H column: 01
Seq primer: T7 primer.
location/Qualifiers
1. .469
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="PBX0018H01"
/clone_lib="Blackshear/Soares normalized Xenopus egg
library"
```

```
library"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"
/note="Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery', Genome Research
6:791-806, 1996. The first strand synthesis used a
NotI-dT18 primer; double stranded cDNAs were ligated to
EcoRI adapters, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested pT7T3-Pac vector.
The library contained approximately 7.2 x 10^5
recombinants, with average insert sizes of 1-1.5 kb."
```

BASE COUNT
ORIGIN

134 a 79 c 136 g 120 t

Query Match 4.3%; Score 20; DB 10; Length 469;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 agagaaagagaatgggacag 23
|||||
Db 88 AGAGAAAGAGATGGGACAG 107

Search completed: April 6, 2002, 05:51:30
Job time: 20662 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2002, 08:22:57 ; Search time 82 Seconds
(without alignments)
1281.534 Million cell updates/sec

Title: US-09-621-781-1_COPY_1507_1970

Perfect score: 464
Sequence: 1 gaaagagaaagagaatggga.....tttgcaaacgcgaaccgcgac 464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_NA:*
2: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81.2	17.5	2344	3	US-08-893-852A-2
2	33.8	7.3	28804	2	US-08-592-874-1
3	33.8	7.3	28804	3	US-09-096-942-2
4	33.8	7.3	28804	3	US-09-096-867-2
5	32.4	7.0	1599	1	US-08-288-405A-9
6	31	6.7	1294	3	US-09-025-691-2
7	30.8	6.6	1718	1	US-08-675-153-3
8	30.8	6.6	1718	2	US-08-841-252-3
9	30.8	6.6	1718	2	US-08-881-571-3
10	30.8	6.6	1718	4	US-09-282-054-3
11	30.8	6.6	1867	1	US-08-675-153-6
12	30.8	6.6	1867	2	US-08-841-252-6
13	30.8	6.6	1867	2	US-08-881-571-6
14	30.8	6.6	1867	4	US-09-282-054-6
15	30.6	6.6	28804	2	US-08-592-874-1
16	30.6	6.6	28804	3	US-09-096-942-2
17	30.6	6.6	28804	3	US-09-096-867-2
18	30.2	6.5	1557	5	PCT-US91-01327-12
19	30.2	6.5	1557	6	5248670-2
20	29.8	6.4	3683	4	US-08-450-962-1
21	29.2	6.3	1785	3	US-08-729-416C-8
22	29	6.2	3363	3	US-09-258-349-1
23	29	6.2	4826	4	US-09-192-983-3
24	29	6.2	43795	3	US-08-742-185-101
25	28.6	6.2	640	3	US-09-154-083-3
26	28.6	6.2	33529	4	US-09-144-085-3
27	28.4	6.1	289	4	US-09-007-005-17

C	28	28.4	6.1	289	4	US-09-244-796-17	Sequence 17, Appl
C	29	28.4	6.1	954	2	US-08-973-275-3	Sequence 3, Appli
C	30	28.4	6.1	1610	4	US-09-276-531-56	Sequence 56, Appl
	31	28.4	6.1	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	32	28.2	6.1	649	4	US-09-040-984-22	Sequence 22, Appl
C	33	28.2	6.1	1449	4	US-09-088-425-4	Sequence 4, Appli
C	34	28	6.0	6611	1	US-08-402-282-2	Sequence 2, Appli
C	35	28	6.0	6611	1	US-08-508-004-2	Sequence 2, Appli
C	36	28	6.0	6611	1	US-08-402-066-2	Sequence 2, Appli
C	37	28	6.0	6611	1	US-08-402-068-2	Sequence 2, Appli
C	38	28	6.0	18073	4	US-09-078-294-12	Sequence 12, Appl
C	39	27.8	6.0	3603	1	US-08-188-582-15	Sequence 15, Appl
C	40	27.8	6.0	3603	1	US-08-646-715-15	Sequence 15, Appl
C	41	27.8	6.0	7218	1	US-08-232-463-14	Sequence 14, Appl
C	42	27.8	6.0	51259	3	US-08-781-891-209	Sequence 209, App
C	43	27.6	5.9	362	1	US-08-486-895-2	Sequence 2, Appli
C	44	27.6	5.9	675	4	US-08-998-416-723	Sequence 723, App
C	45	27.6	5.9	1715	4	US-08-757-230A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-893-852A-2
; Sequence 2, Application US/08893852A
; Patent No. 6080558
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,852A
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0341 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2344 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TMLR3DT01
; CLONE: 508302
; US-08-893-852A-2

Query Match 17.5%; Score 81.2; DB 3; Length 2344;
Best Local Similarity 70.9%; Pred. No. 5.4e-16;

Matches 122; Conservative 0; Mismatches 48; Indels 2; Gaps 1;

QY 293 gtggaagattacatgagagaccccgcgactccgcatcccttgcgggaagaccttg 352

Db 8 GTTGTGATCTTATGCAGACGCTGCACGACCCCGCCCGCTGTGCGCCAGGCACTTG 67

QY 353 cgacagcccgtagacatcaagt--ccccagcccccaagcctgagggagacatgaagcg 410

Db 68 AGGCAGCCGGAGATACTGTGAGTTACTCGGAGCCCGACGCTGAGGGTGAAGATGAACGCG 127

QY 411 ctggccttgagagcaatcccgagaccagatcgcttggcaaacgaacgg 462

Db 128 CTGGCTCCCTAACCGTCCGGAACCTGTGATCGCTTGTGCGAGACCGAACCGG 179

RESULT 2

US-08-592-874-1

; Sequence 1, Application US/08592874

; Patent No. 5854034

; GENERAL INFORMATION:

; APPLICANT: POLLOCK, THOMAS J.

; APPLICANT: YAMAZAKI, MOTOHIDE

; APPLICANT: THORNE, LINDA

; APPLICANT: MIKOLAJCZAK, MARCIA

; APPLICANT: ARMENTROUT, RICHARD W.

; TITLE OF INVENTION: DNA SEGMENTS AND METHODS FOR INCREASING

; TITLE OF INVENTION: POLYSACCHARIDE PRODUCTION

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: JULES E. GOLDBERG

; STREET: 261 MADISON AVENUE

; CITY: NEW YORK

; STATE: NY

; COUNTRY: USA

; ZIP: 10016-2391

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/592,874

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/377,440

; FILING DATE: 24-JAN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: GOLDBERG, JULES E.

; REGISTRATION NUMBER: 24,408

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-986-4090

; TELEFAX: 212-818-9479

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 28804 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: unknown

; TOPOLOGY: unknown

; MOLECULE TYPE: DNA (genomic)

; FRAGMENT TYPE: N-terminal

; US-08-592-874-1

Query Match

Best Local Similarity 7.3%; Score 33.8; DB 2; Length 28804;

Matches 80; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 301 ttacatgcgagaccccgcgactccgcatcccttgcgggaagacgcttgagacagcc 360

Db 24702 TTACCGCGCGCGCGCTTCATCGGCTCGGAGTGTGACGCCACCTCGTTGCGCAGGGCG 24761

QY 361 cgtgagacatcacgtccccgagcccccaagcctgagggcgacatgaacgcgtgcttg 420

Db 24762 CGCGCGTCATCATCTCGACACAGCTCACCTATGCGGGCAACCCGGCTCGCTGACCGCA 24821

QY 421 gagcaatccgagaccacagatcgcttggcaaacgga 457

Db 24822 TCGAGACGCCCCCACTACTACCGCTTGTGTCACGCGCA 24858

RESULT 3

US-09-096-942-2

; Sequence 2, Application US/09096942

; Patent No. 6027925

; GENERAL INFORMATION:

; APPLICANT: POLLOCK, Thomas J

; APPLICANT: Mikolajczak, Marcia

; APPLICANT: Yamazaki, Motohide

; APPLICANT: Thorne, Linda

; APPLICANT: Armentrout, Richard W

; TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria

; TITLE OF INVENTION: Carrying Genes from Xanthomonas Campestris

; FILE REFERENCE: seq list for appl filed from pro. appl

; CURRENT APPLICATION NUMBER: US/09/096,942

; EARLIER FILING DATE: 1998-06-12

; EARLIER APPLICATION NUMBER: 60/049,428

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 2

; LENGTH: 28804

; TYPE: DNA

; ORGANISM: Sphingomonas sp. S88

; US-09-096-942-2

Query Match

Best Local Similarity 7.3%; Score 33.8; DB 3; Length 28804;

Matches 80; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 301 ttacatgcgagaccccgcgactccgcatcccttgcgggaagacgcttgagacagcc 360

Db 24702 ttaccgcgcgcgcgcttcacatcgctcgcgagtgtagccacatcgcttcgcaagcg 24761

QY 361 cgtgagacatcacgtccccgagcccccaagcctgagggcgacatgaacgcgtgcttg 420

Db 24762 cgcgcgtcatcatctcgacaagctcacctatgcggcaacccgcgtcgtgacgcgca 24821

QY 421 gagcaatccgagaccacagatcgcttggcaaacgga 457

Db 24822 tcgagaacgcccccaactaccgcttcgtccacgcgca 24858

RESULT 4

US-09-096-867-2

; Sequence 2, Application US/09096867

; Patent No. 6030817

; GENERAL INFORMATION:

; APPLICANT: Pollock, Thomas J

; APPLICANT: Mikolajczak, Marcia

; APPLICANT: Yamazaki, Motohide

; APPLICANT: Thorne, Linda

; APPLICANT: Armentrout, Richard W

; TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria

; TITLE OF INVENTION: Carrying Genes from Xanthomonas Campestris

; FILE REFERENCE: seq list for appl filed from pro. appl

; CURRENT APPLICATION NUMBER: US/09/096,867

; EARLIER FILING DATE: 1998-06-11

; EARLIER APPLICATION NUMBER: 60/049,428

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 2

; LENGTH: 28804

; TYPE: DNA


```
RESULT 7
US-08-675-153-3/c
; Sequence 3, Application US/08675153
; Patent No. 5677124
; GENERAL INFORMATION:
; APPLICANT: Dubois, Dwight
; APPLICANT: Winkler, Matthew
; APPLICANT: Pasloske, Brittan L.
; TITLE OF INVENTION: RIBONUCLEASE RESISTANT VIRAL
; TITLE OF INVENTION: RNA STANDARDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/675,153
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: AMBI:026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1718 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-675-153-3

Query Match 6.6%; Score 30.8; DB 1; Length 1718;
Best Local Similarity 48.3%; Pred. No. 2.8;
Matches 86; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 136 ccggaatctccgcgagaatttcagtggtgttctcctcctccaccttctcagggacttc 195
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1495 CAGGAAGCTCTACACCAACACAGCTCTGGGTTGCCACTTTAGGCACTCGACTTGTATGG 1436

QY 196 cgaactccgcctctccggtgacgtcagcatagcgtcgtcagactataaactccggg 255
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1435 TGTATTTCGATTCTGCGCAGAGCTCTGACGAACGCTACAGGTTACTTTGTAAGCCTGTG 1376

QY 256 tgatcgtgttgccgagattgactcagttcgcaagcttggtgaagattacatgcgagac 313
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1375 AACGCGAGTTAGAGCTGATTCATTCAAGCAGCCCGTTAGCGAAGTTGCTTGGGGCGAC 1318

RESULT 8
US-08-841-252-3/c
; Sequence 3, Application US/08841252
; Patent No. 5919625
; GENERAL INFORMATION:
; APPLICANT: DUBOIS, DWIGHT
; APPLICANT: WINKLER, MATTHEW
; APPLICANT: PASLOSKE, BRITTAN L.
; TITLE OF INVENTION: RIBONUCLEASE RESISTANT VIRAL RNA
; TITLE OF INVENTION: STANDARDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD WHITE & DURKEE
```

```
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/841,252
; FILING DATE: 29-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 5,677,124
; FILING DATE: 03-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARK B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: AMBI:026--1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-300
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1718 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-841-252-3

Query Match 6.6%; Score 30.8; DB 2; Length 1718;
Best Local Similarity 48.3%; Pred. No. 2.8;
Matches 86; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 136 ccggaatctccgcgagaatttcagtggtgttctcctcctccaccttctcagggacttc 195
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1495 CAGGAAGCTCTACACCAACACAGCTCTGGGTTGCCACTTTAGGCACTCGACTTGTATGG 1436

QY 196 cgaactccgcctctccggtgacgtcagcatagcgtcgtcagactataaactccggg 255
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1435 TGTATTTCGATTCTGCGCAGAGCTCTGACGAACGCTACAGGTTACTTTGTAAGCCTGTG 1376

QY 256 tgatcgtgttgccgagattgactcagttcgcaagcttggtgaagattacatgcgagac 313
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1375 AACGCGAGTTAGAGCTGATTCATTCAAGCAGCCCGTTAGCGAAGTTGCTTGGGGCGAC 1318

RESULT 9
US-08-881-571-3/c
; Sequence 3, Application US/08881571
; Patent No. 5939262
; GENERAL INFORMATION:
; APPLICANT: Pasloske, Brittan L.
; APPLICANT: Dubois, Dwight
; APPLICANT: Brown, David
; APPLICANT: Winkler, Matthew
; TITLE OF INVENTION: RIBONUCLEASE RESISTANT RNA PREPARATION
; TITLE OF INVENTION: AND UTILIZATION
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/881,571
: FILING DATE: Concurrently Herewith
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/675,153
: FILING DATE: 03-JUL-1996
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/021,145
: FILING DATE: 03-JUL-1996
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Wilson, Mark B.
: REGISTRATION NUMBER: 37,259
: REFERENCE/DOCKET NUMBER: AMBI:033
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 512/418-3000
: TELEFAX: 512/474-7577
:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1718 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-08-881-571-3

```

Query Match	6.68;	Score 30.8;	DB 2;	Length 1718;
Best Local Similarity	48.38;	Pred. No. 2.8;		
Matches	86;	Conservative	0;	Mismatches 92;
			Indels	0;
			Gaps	0;

QY	136	ccggaatctccgcgagaatttcagtggtgtgttccctctctccaccttctcagygacttc	195
Db	1495	caggaagctctacacaccacaacagctctgggttgccacttttaggcaccctgcactttgatgg	1436
QY	196	cgaactccgcctctccggtgacgtcagcatagcgtcgtcagactataaactccggg	255
Db	1435	tgtatttggcattctgcgcagagctctgacgaacgctacaggttactttgtaagcctcttg	1376
QY	256	tgatcgtgttgcgcagattgactcagltcgcagcttctggaagattacatgcgagac	313
Db	1375	aacgcgagttagagctgattccattcagcgcacccccgttacgcaagttgctttggggccgac	1318

```

RESULT 10
US-09-282-054-3/c
; Sequence 3, Application US/09282054
; Patent No. 6214982
;
; GENERAL INFORMATION:
;
; APPLICANT: Pasloske, Brittan L.
; APPLICANT: Dubois, Dwight
; APPLICANT: Brown, David
; APPLICANT: Winkler, Mathew
;
; TITLE OF INVENTION: RIBONUCLEASE RESISTANT RNA PREPARATION
; TITLE OF INVENTION: AND UTILIZATION
;
; NUMBER OF SEQUENCES: 8
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
;
; STATE: Texas
; COUNTRY: USA
;
; ZIP: 77210
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/282,054
; FILING DATE:
; CLASSIFICATION:
;
; PRIOR APPLICATION DATA:

```

```

: APPLICATION NUMBER: US/08/881,571
: FILING DATE:
: APPLICATION NUMBER: US 08/675,153
: FILING DATE: 03-JUL-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/021,145
: FILING DATE: 03-JUL-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Wilson, Mark B.
: REGISTRATION NUMBER: 37,259
: REFERENCE/DOCKET NUMBER: AMB1:033
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 512/418-3000
: TELEFAX: 512/474-7577
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1718 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
US-09-282-033

```

Query Match	6.6%;	Score 30.8;	DB 4;	Length 1718;
Best.Local Similarity	48.3%;	Pred. No. 2.8;		
Matches	86; Conservative	0; Mismatches	92; Indels	0; Gaps

QY	136	ccggaatctccgcgagaatttcagltgtgtttctctctccaccttctcagggacttc	195
Db	1495	CAGGAAGCTCTACACACCACCAACAGTCTGGGTGGCAGCTTAAGGCACCTCGACTTTGATGG	1436
QY	196	cgaactccgcctctccggtgacgtcagcatagcgtcgtcagactataaactccggy	255
Db	1435	TGTAATTGGGATTTCTGCGCAGAGCTCTGACGAAACGCTACAGTTACTTTGTAAGCCTGTG	1376
QY	256	tgatcgtgttgcgcagattgaactcagltcgcagcttgtggaagattacatgcyagac	313
Db	1375	AACCGGAGTTAGAGCTGATTCATTCAGCGACCCCGTTAGCGAAGTTGCTTGGGGCGAC	1318

RESULT 11
US-08-675-153-6/c
; Sequence 6, Application US/08675153
; Patent No. 5677124
; GENERAL INFORMATION:
; APPLICANT: Dubois, Dwight
; APPLICANT: Winkler, Matthew
; APPLICANT: Pasloske, Brittan L.
; TITLE OF INVENTION: RIBONUCLEASE RESISTANT VIRAL
; TITLE OF INVENTION: RNA STANDARDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/675,153
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: AMBI:026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000

QY	Db	QY	Db
196	1495	196	1495
cgaactccgcctctccggtgacgyltcagcatagcgtcgltcagactataactccggg	CAGGAAGCTCTACACCACCACAACAGCTCTGGGTTGCCACTTTAGGCACCTGCACCTTGTATGG	cgaactccgcctctccggtgacgyltcagcatagcgtcgltcagactataactccggg	CAGGAAGCTCTACACCACCACAACAGCTCTGGGTTGCCACTTTAGGCACCTGCACCTTGTATGG
1435	1375	1435	1375
TGTATTTGGCATTTCTGGCGCAGAGCTCTGACGAACCGCTACAGGTTACTTTGTAAGCCCTGTG	AACGGCGAGTTAGAGCTGTATCCATTTCAGCGACCCCCGTTAGCGAAGTTGCTTTGGGGCGAC	TGTATTTGGCATTTCTGGCGCAGAGCTCTGACGAACCGCTACAGGTTACTTTGTAAGCCCTGTG	AACGGCGAGTTAGAGCTGTATCCATTTCAGCGACCCCCGTTAGCGAAGTTGCTTTGGGGCGAC
256	1318	256	1318
tgatcgtgttgcgcgagatttactcagttcgcagcttgttggaaattacatgcgagac	1318	tgatcgtgttgcgcgagatttactcagttcgcagcttgttggaaattacatgcgagac	1318

RESULT 14

```

US-09-282-054-6/c
; Sequence 6, Application US/09282054
; Patent No. 6214982
; GENERAL INFORMATION:
; APPLICANT: Pasloske, Brittan L.
; APPLICANT: Dubois, Dwight
; APPLICANT: Brown, David
; APPLICANT: Winkler, Matthew
; TITLE OF INVENTION: RIBONUCLEASE RESISTANT RNA PREPARATION
; TITLE OF INVENTION: AND UTILIZATION
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; *COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/282,054
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/881,571
; FILING DATE:
; APPLICATION NUMBER: US 08/675,153
; FILING DATE: 03-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/021,145
; FILING DATE: 03-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: AMBI:033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1867 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-282-054-6

Query Match          6.6%; Score 30.8; DB 4; Length 1867;
Best Local Similarity 48.3%; Pred. No. 2.9;
Matches 86; Conservative 0; Mismatches 92; Indels 0; Gaps

QY 136 ccgaatctccgcgcgagaattcaagtgtgtttcctctccacacttctcagygacttc 195
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1495 CAGGAAGCTCTACACACCAACACAGACTGTGGGTTCGCCACTTTAGGCACCTCGACTTTGATGG 1436

QY 196 cgaactccgcctctccggtgacgtcagcatagcgctgcgtcagactataaactcccg 255
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

[illegible]

RESULT 15

```

US-08-592-874-1/c
; Sequence 1, Application US/08592874
; Patent No. 5854034
;
; GENERAL INFORMATION:
; APPLICANT: POLLOCK, THOMAS J.
; APPLICANT: YAMAZAKI, MOTOHIDE
; APPLICANT: THORNE, LINDA
; APPLICANT: MIKOLAJCZAK, MARCIA
; APPLICANT: ARMENTROUT, RICHARD W.
; TITLE OF INVENTION: DNA SEGMENTS AND METHODS FOR INCREASING
; TITLE OF INVENTION: POLYSACCHARIDE PRODUCTION
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JULES E. GOLDBERG
; STREET: 261 MADISON AVENUE
; CITY: NEW YORK
; STATE: NY
;
; COUNTRY: USA
; ZIP: 10016-2391
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,874
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/377,440
; FILING DATE: 24-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: GOLDBERG, JULES E.
; REGISTRATION NUMBER: 24,408
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-986-4090
; TELEFAX: 212-818-9479
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28804 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FRAGMENT TYPE: N-terminal
;
US-08-592-874-1

```

Query Match	6.6%;	Score 30.6;	DB 2;	Length 28804;
Best Local Similarity	49.1%;	Pred. No. 10;		
Matches 81; Conservative	0;	Mismatches 84;	Indels 0;	Gaps 0;

[illegible]

Sat Apr 6 11:03:28 2002

us-09-621-781-1_copy_1507_1970_1.rni

Page 8

Search completed: April 5, 2002, 08:48:45
Job time: 1548 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2002, 08:22:57 ; Search time 1414.97 Seconds
(without alignments)
3523.779 Million cell updates/sec

Title: US-09-621-781-1_COPY_1507_1970

Perfect score: 464
Sequence: 1 gaaagagaagaagaatgga.....tttgcaaacggaaccgac 464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_est1:*
11: gb_est2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inh:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	155	33.4	623	10	AW320574	AW320574 uc20f09.y
2	135.4	29.2	575	10	AA472295	AA472295 vh01c05.r
3	135.4	29.2	615	11	W62865	W62865 md85f08.r1
4	134	28.9	542	10	AW762360	AW762360 ur56h03.y
5	121.6	26.2	573	10	AA104251	AA104251 mo50b03.r
6	114.6	24.7	461	10	AA920479	AA920479 vy53c08.r
7	111.8	24.1	908	11	BF182426	BF182426 601804187
8	108	23.3	708	10	BE573350	BE573350 601333015
9	107.8	23.2	596	10	BE573255	BE573255 601333152
c 10	103.8	22.4	446	10	AA684447	AA684447 vm64b07.s
11	96	20.7	1136	10	BE573733	BE573733 601333482
12	86.4	18.6	574	11	BG100299	BG100299 ux84f03.y

13	84	18.1	778	11	BG535582	BG535582 602563276
14	82.8	17.8	773	11	BF106314	BF106314 601823528
15	81.2	17.5	642	10	AU126992	AU126992 AU126992
16	81.2	17.5	714	11	BG621191	BG621191 602616890
17	81.2	17.5	714	11	BF311344	BF311344 601896744
18	81.2	17.5	719	10	AU137751	AU137751 AU137751
19	81.2	17.5	747	2	BG752598	Bg752598 602731110
20	81.2	17.5	882	11	BI260970	BI260970 602972125
21	80.2	17.3	411	10	AU136946	AU136946 AU136946
22	80.2	17.3	864	11	BF238180	BF238180 601811913
23	80	17.2	558	11	BG478190	BG478190 602522850
24	80	17.2	721	10	AL134765	AL134765 DKFZP547G
25	79.6	17.2	516	10	AU128763	AU128763 AU128763
26	79.6	17.2	901	10	AL554828	AL554828 AL554828
27	78	16.8	383	10	BE246464	BE246464 TCBAPIE47
28	75.8	16.3	836	10	BE312726	BE312726 601149690
29	75.4	16.3	186	10	AA381292	AA381292 EST94363
30	75.4	16.3	1035	10	BE613683	BE613683 601504479
31	75.4	16.3	1068	11	BF037888	BF037888 601461435
32	74.4	16.0	927	10	AL543823	AL543823 AL543823
33	72.2	15.6	752	11	BI148394	BI148394 602914248
34	71.4	15.4	613	11	BG476937	BG476937 602522145
35	70.2	15.1	749	11	BG500419	BG500419 602544931
36	69.4	15.0	430	11	R93673	R93673 yq34e06.r1
37	68.4	14.7	872	10	BE253631	BE253631 601109077
38	67.2	14.5	300	10	AU098820	AU098820 AU098820
39	66.8	14.4	306	10	AU076649	AU076649 AU076649
40	66.8	14.4	695	11	BI224257	BI224257 602940548
41	65.8	14.2	543	11	BF128750	BF128750 601811129
42	65.4	14.1	936	11	BI113465	BI113465 602899696
43	64.4	13.9	476	11	BG394934	BG394934 602457321
44	63.6	13.7	592	10	BE313235	BE313235 601151521
45	59.2	12.8	1261	11	BG425501	BG425501 602447367

ALIGNMENTS

RESULT 1
LOCUS AW320574 623 bp mRNA EST 25-JAN-2000
DEFINITION uc20f09.y1 NCI-CGAP_Mam6 Mus musculus cDNA clone IMAGE:2609705 5' similar to gb:X51829 Mouse myeloid differentiation primary response mRNA encoding (MOUSE);, mRNA sequence.

ACCESSION AW320574
VERSION AW320574.1 GI:6750118

KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 623)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/bbrp/image/image.html

FEATURES
source MGI:1023565
Seq primer: -40RP from Gibco
High quality sequence stop: 434.
Location/Qualifiers
1. .623
/organism="Mus musculus"
/strain="FVB/N"

```
/db_xref="taxon:10090"
/clone="IMAGE:2609705"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/Note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      137 a      197 c      173 g      115 t      1 others
ORIGIN
```

```
Query Match      33.4%; Score 155; DB 10; Length 623;
Best Local Similarity 88.3%; Pred. No. 2.7e-34;
Matches 181; Conservative 0; Mismatches 20; Indels 4; Gaps 1;
```

```
Y      262  tgttgccgagattgactcagtcgcagctgtggaagattacatgcgagaccgcgcg 321
      4  TGTGGCGCAGATTGAGTCAGCTGTGAGTTGTGGAAGATTACATGCATATCCCGCGCG 63
QY     322  actccgcacccctt----gccggagacagccttgcgcagaccgctgagacatcacgtcc 377
      64  ACCCCGCATCCCTTTGCCGCGCGGACAGCCTTTGCTACAGCCTGTGAACATTTGCCGTCC 123
QY     378  ccgagcccccacgcctgagggcgacatgaacgcgctgccttgagagacaatccgagaccac 437
      Db  124  CCGAGCCCCACGCCCTGAGGGCGACATGAACCCGCTGGCTTCGCCAGACAGTCCGACCCAC 183
QY     438  gatcgcttttgcaaacgacgcg 462
      Db  184  GATCGCTTTTGCAACACGAAACCGG 208
```

RESULT 2
AA472295
LOCUS AA472295 575 bp mRNA EST 18-JUN-1997
DEFINITION vhlc05.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
IMAGE:874184 5' similar to gb:X51829 Mouse myeloid differentiation
primary response mRNA encoding (MOUSE);, mRNA sequence.
ACCESSION AA472295
VERSION AA472295.1 GI:2200286
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 575)
REFERENCE Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:513664
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 377.
Location/Qualifiers
1..575
/Organism="Mus musculus"
/strain="C57BL/6J"

FEATURES

source

```
/db_xref="taxon:10090"
/clone="IMAGE:874184"
/clone_lib="Soares_mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/Note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5'
TGTACCAATCTGAAGTGGAGCGCGCCGCGAATGGTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT      127 a      183 c      161 g      104 t
ORIGIN
```

```
Query Match      29.2%; Score 135.4; DB 10; Length 575;
Best Local Similarity 88.8%; Pred. No. 1.3e-28;
Matches 159; Conservative 0; Mismatches 16; Indels 4; Gaps 1;
```

```
QY     288  agcttgtggaagattacatgcgagaccgcgcgactccgcacccctt----gccggga 343
      Db  4  AGTTGTGGAAGATTACATGCATATGCCGCGGACGCCGCGCATCCCTTTGCCGCGCGGGA 63
QY     344  cagccttgcgacagcccgctgagacatcacgtlcccgagcccaacgcctgagggcgacat 403
      Db  64  CAGCCTTGTGCTACAGCCTGTGAACAATTCGTCGCCGAGCCCAACGCTTGAGGGCGACAT 123
QY     404  gaacgcgtgcgcttgagagacaatccgagaccacagatcgctttggcaaacgacgcg 462
      Db  124  GAACCCGCTGGCTTCGCGAGCAGTCCGACCCACGATCGCTTTTGCAACACGAAACCGG 182
```

RESULT 3
W62865
LOCUS W62865 615 bp mRNA EST 07-JUN-1996
DEFINITION md85f08.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
clone IMAGE:375207 5' similar to gb:X51829 Mouse myeloid
differentiation primary response mRNA encoding (MOUSE);, mRNA
sequence.
ACCESSION W62865
VERSION W62865.1 GI:1369614
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 615)
REFERENCE Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:236639
Seq primer: ETPrimer
High quality sequence stop: 344.

TITLE
JOURNAL
COMMENT

/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

BASE COUNT 220 a 271 c 264 g 153 t
ORIGIN

Query Match 24.1%; Score 111.8; DB 11; Length 908;
Best Local Similarity 85.1%; Pred. No. 9.3e-22;
Matches 149; Conservative 0; Mismatches 22; Indels 4; Gaps 2;

QY 291 ttgtggaagattacatgcgagaccgcgcgactccgcatccctt--tgccggagacagc 347
|||||
Db 7 TTGTGGAAGATTACATGCGATATCCCGCGAGCCGCATCCCTTTGCCGCGCGGACAGC 66

QY 348 ctttgcgacagcccgtagacatcacgtcccccagagcccccagcctgagggcgacatgaac 407
|||||
Db 67 CTTTGCTACAGCCTGTGAACATGTGCGTCCCGAGCCCGACGCTGAGGGCGACATGAAC 126

QY 408 ggcgtgacctgagagcaatcccgagaccacagatcgctttggcaaacggaacgg 462
|||||
Db 127 CCGCTGGCTTCGCGAGCAGTCCGAGCCACGATCGC-TTTGGCAACGAGAACCGG 180

RESULT 8
BE573350 708 bp mRNA EST 15-AUG-2000
LOCUS 601333015F2 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3710762 5',
DEFINITION mRNA sequence.

ACCESSION BE573350
VERSION BE573350.1 GI:9817070
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 708)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM8933 row: m column: 03
High quality sequence stop: 649.

FEATURES
source location/Qualifiers
1..708
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3710762"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 165 a 213 c 208 g 122 t
ORIGIN

Query Match 23.3%; Score 108; DB 10; Length 708;
Best Local Similarity 87.7%; Pred. No. 1.1e-20;
Matches 142; Conservative 0; Mismatches 15; Indels 5; Gaps 2;

QY 305 atcgagaccgcgcgactccgcatccctt---gccggagacgcttgcgacagcc 360
|||||
Db 1 ATGCGATATCCCGCGAGACCCCGCATCCCTTTGCCGCGCGGACAGCCTTTGCTACAGCC 60

QY 361 cgtgagacatcacgtcccccagagcccccagcctgagggcgacatgaacgcgtgacctga 420
|||||
Db 61 TGTGAACATTTGCGTCCCGAGCCCGACGCTGAGGGCGACATGAACCCGCTGCTTCGC 120

QY 421 gagcaatccgagaccacagatcgctttggcaaacggaacgg 462
|||||
Db 121 GAGCAGTCCGAGCCACGATCGC-TTTGGCAACGAGAACCGG 161

RESULT 9

BE573255 596 bp mRNA EST 15-AUG-2000
LOCUS 601333152F2 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3710629 5',
DEFINITION mRNA sequence.

ACCESSION BE573255
VERSION BE573255.1 GI:9816975
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 596)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM8933 row: g column: 14
High quality sequence stop: 542.

FEATURES
source location/Qualifiers
1..596
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3710629"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 132 a 199 c 163 g 102 t
ORIGIN

Query Match 23.2%; Score 107.8; DB 10; Length 596;
Best Local Similarity 84.8%; Pred. No. 1.2e-20;
Matches 145; Conservative 0; Mismatches 22; Indels 4; Gaps 2;

QY 295 ggaagattacatgcgagaccgcgcgactccgcatccctt--tgccggagacgctt 351
|||||
Db 1 GGAAGATTACATGCGATATCCCGCGAGCCCGCATCCCTTTGCCGCGCGGAGACAGCCTTT 60

QY 352 ggcacagcccgtagacatcacgtcccccagagcccccagcctgagggcgacatgaacggc 411
|||||

ACCESSION	BG100299
VERSION	BG100299.1 GI:12595616
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 574)
AUTHORS	Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R. The WashU-NCI Mouse EST Project 1999
TITLE	Unpublished (1999)
JOURNAL	Contact: Marra M/WashU-NCI Mouse EST Project 1999
COMMENT	Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MG1:1417724 Seq primer: Primer name ambiguous High quality sequence stop: 392.
FEATURES	Location/Qualifiers
source	1..574
	/organism="Mus musculus"
	/strain="CD-1"
	/db_xref="taxon:10090"
	/clone="IMAGE:3655420"
	/clone_lib="McCarrey Eddy type B spermatogonia"
	/sex="male"
	/tissue_type="type B spermatogonia, pooled from multiple mice"
	/dev_stage="8 day"
	/lab_host="DH10B (phage-resistant)"
	/note="Organ: testis; Vector: pBluescript SK+ (Stratagene)"; Site_1: XhoI; Site_2: EcoRI; cDNA oligo dT-primed [5'-(GA)10-ACTAGTCGTGGAGTTTATTTTTTTT-3'] and directionally cloned using 5' linkers 5'-AATTGGCAGCAG-3' and 5'-CTCGTGCCG-3'. Size selection of >400bp material gives average insert size ranging from 1-2 kb. Library was mass excised (from lambda-UniZAP-XR) and resulting single-stranded phagemids were prepped and transformed into DH10B. Library contains 96% recombinants. References: J. Androl. 20:635-639 and Gene 25:263-269. Library constructed and donated by J. McCarrey, Ph.D. (Southwest Foundation for Biomedical Research, Dept. of Genetics); excision done by F.M. Eddy, Ph.D. (National Institutes of Health, National Institute of Environmental Health Sciences). Original lambda-based library is available through ATCC, catalog #63417."
BASE COUNT	133 a 179 c 159 g 100 t 3 others
ORIGIN	.
Query Match	18.6%; Score 86.4; DB 11; Length 574;
Best Local Similarity	89.4%; Pred. No. 1.8e-14;
Matches	93; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
OY	359 ccocgtgacatcacgtcccccaggccccaccgcctgaggcgacatgaacgcgtgccct 418
Dd	37 CCTGTGAACAATTTGCCTCCCAGAGCCCCACGCCCTGAGGGCAATGAACCCGCTGCTTC 96
OY	419 gagagcaatcggaccacgcatcgcttttgccaacccgaacgy 462
Dd	97 GCGAGCATGTCGAGCACCATCGCTTTTGCAACCAGAACC GG 140

RESULT 13
BG535582

LOCUS	BG535582	778 bp	mRNA	EST	03-APR-2001
DEFINITION	602563276F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4687939 5', mRNA sequence.				
ACCESSION	BG535582				
VERSION	BG535582.1	GI:13527127			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 778)				
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)				
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: CLONTECH Laboratories, Inc. CDNA Library Preparation: CLONTECH Laboratories, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLCML501 row: h column: 20 High quality sequence stop: 510.				
FEATURES	Location/Qualifiers 1..778 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4687939" /clone_lib="NIH_MGC_77" /lab_host="DH10B (TI phage-resistant)" /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1: sfil (ggccgcctcgcc); Site_2: sfil (ggccatatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:: 5'-ATTCTAGAGGCCCGAGGGCGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC library."				
BASE COUNT	166 a	239 c	249 g	124 t	
ORIGIN					
Query Match	18.1%;	Score 84;	DB 11;	Length 778;	
Best Local Similarity	69.4%;	Pred. No. 9.8e-14;			
Matches 129;	Conservative 0;	Mismatches 55;	Indels 2;	Gaps 1;	
OY	279	tcagttcgca	gcttgtgtga	aatcatcat	gcgagacccccgcgcgactccgcattcgc 338
Db	9	TCGGTTCCCA	CTCCCACTG	GTGAATCTTA	TGCAAGACGCTGCACAGACCCCCGGCCGCTTG 68
OY	339	cgggacacg	cctttgcg	acagcccagt	gagacatcacgt--ccccgaagccccacgcctgagg 396
Db	69	CGCCACGGCA	CTTGAGGCA	GCGCAGACTA	CTCTGAGTTACTCGGAGCCCCGACGCGCTGAGG 128
OY	397	gcgacatga	acgcgcctg	gccttgaga	gcaatccgcgaccacaagatcgttttggcaaacg 456
Db	129	GTCGAGATGA	ACGCGCTGG	CGCTCCCTTA	ACCCTCCGACCTGTGATGCTTCTTGCGACAGCCG 188
OY	457	aaccgg	462		
Db	189	AACCGG	194		
RESULT 14					
LOCUS	BF106314	773 bp	mRNA	EST	19-OCT-2000
DEFINITION	601823528F1 NIH_MGC_77 Homo sapiens CDNA clone IMAGE:4043190 5', mRNA sequence.				
ACCESSION	BF106314				

VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
BF106314.1	GI:10888840	EST.	human.	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	1 (bases 1 to 773)	NIH-MGC	http://mgs.nci.nih.gov/.	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
				Contact: Robert Strausberg, Ph.D.				Email: cgapbs-remail.nih.gov
				Tissue Procurement: CLONTECH Laboratories, Inc.				CDNA Library Preparation: CLONTECH Laboratories, Inc.
				CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)				DNA Sequencing by: Incyte Genomics, Inc.
				Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:				http://image.lnl.gov
				Plate: LICM865	row: h	column: 07		High quality sequence stop: 657.
				location/Qualifiers				
				1. .773				
				/organism="Homo sapiens"				
				/db_xref="taxon:9606"				
				/clone="IMAGE:4043190"				
				/clone_lib="NIH_MGC_77"				
				/lab_host="DH10B (T1 phage-resistant)"				
				/note="Organ: Lung; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgctggcc); Site_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCCGAGCGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."				
				BASE COUNT				
				162 a	232 c	243 g	136 t	
				ORIGIN				
				Query:Match	17.8%;	Score 82.8;	DB 11;	Length 773;
				Best Local Similarity	71.5%;	Pred. No. 2.2e-13;		
				Matches 123;	Conservative 0;	Mismatches 47;	Indels 2;	Gaps 1;
				QY 293	gtggaagattacatgcgagaccccgcgactccgcatcccttgcgagacgcttg	352		
				1111 111 1111 1111 1111 1111 111 111 111 111				
				4 GTGGTTGATCTTATGCAGACGCTGCACGACCCCGCGCGCTTGTGCGCACGGCATTG	63			
				QY 353	cgacagcccgtagacatcacgt--ccccgagcccccaagcctgagggcgacatgaacgcg	410		
				11111 111 111 111 111 11111 111111111 111111111				
				Db 64	AGGCAGCCGAGATACCTGAGTTACTCGGAGCCCGACGCGCTGAGGGTGAGATGAACGCG	123		
				QY 411	ctgagccttagagcaatccgagcccccaagcgtcgttcttgcaaacggaacgcg	462		
				1111111 111 1111111 1111111 1111 111111111				
				Db 124	CTGGCCCTCCCTAACCGTCCGAGACCTGTGATCGCTTCTTGCGACGACCGAACCGG	175		
				RESULT 15				
				AU126992				
				LOCUS	AU126992	642 bp	mRNA	EST
				DEFINITION	AU126992	NT2RP2 Homo sapiens	CDNA clone	NT2RP2000346 5', mRNA
				sequence.				
				ACCESSION	AU126992			
				VERSION	AU126992.1	GI:10951708		
				KEYWORDS	EST.			
				SOURCE	human.			
				ORGANISM	Homo sapiens			
					Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
				REFERENCE	1 (bases 1 to 642)			
				AUTHORS	Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,			

TITLE	Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.					
JOURNAL	HRI human cDNA project Unpublished (2000)					
COMMENT	Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3951 Fax: 81-438-52-3952 Email: genomics@hri.co.jp HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute. Location/Qualifiers 1..642					
FEATURES	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="NT2RP200346" /clone_1ib="NT2RP2" /cell_type="teratocarcinoma" /cell_line="NT2" /note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"					
BASE COUNT	131	a	200	c	190	g 115 t 6 others
ORIGIN						
Query Match	17.5%;	Score	81.2;	DB	10;	Length 642;
Best Local Similarity	70.9%;	Pred. No.	6e-13;			
Matches	122;	Conservative	0;	Mismatches	48;	Indels 2; Gaps 1;
QY	293	gtggaagattacatgcgagaaccgccggcgactccgcatttgcgggacagccttg	352			
Dd	23	GTTGTGATCTTATGCAGACGCGTCACGACC CGCCCGCTTGTCGCCACGGCACTTG	82			
QY	353	cgacagcccgtgagacatacgt--ccccgagccccacgcctgaggcgacataaacgcg	410			
Dd	83	AGGCAGCCGGAGATACTCTGAAGTTACTCGGAGGCCCGACCCCTGAGGGTGAGATGAACGG	142			
QY	411	ctggccttagagacaatccgcgaccacacgatcgctttggcaaacgaaccg	462			
Dd	143	CTGGCCTCCCTAACCGCTCCGACCGCTGTGATCGCTTCTGGCAGACCGAACC GG	194			

Search completed: April 5, 2002, 08:46:52
Job time: 1435 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2002, 08:22:57 ; Search time 155.6 seconds
(without alignments)
2556.548 Million cell updates/sec

Title: US-09-621-781-1_COPY_1507_1970
Perfect score: 464
Sequence: 1 gaaagagaagaagaatggga.....tttgcaaacgcgaaccggac 464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
1	428	92.2	2614	19	AAV65766
2	428	92.2	2614	20	AAZ23030
3	131.4	28.3	2111	19	AAV65765
4	131.4	28.3	2111	20	AAZ23029
5	131.4	28.3	2111	20	AAV87383
6	81.8	17.6	287	21	AAC09789
7	81.2	17.5	642	22	AAH05790
8	81.2	17.5	2331	21	AAZ99766
9	81.2	17.5	2344	20	AAV19798
10	81.2	17.5	2349	22	AAH14216
11	37	8.0	1566	21	AAC47239

C	12	37	8.0	1606	21	AAC41089
	13	36.2	7.8	494	20	AAV86970
	14	33.8	7.3	28804	17	AAV37329
	15	33.8	7.3	28804	18	AAV92474
	16	33.8	7.3	28804	20	AAV99812
	17	33.8	7.3	28804	20	AAV81474
	18	33.2	7.2	5160	21	AAV21033
	19	33.2	7.2	5160	21	AAA34911
	20	33.2	7.2	13548	21	AAV21037
	21	33.2	7.2	13548	21	AAA34915
	22	32.4	7.0	1598	16	AAV04953
	23	31.8	6.9	3747	22	AAV26918
	24	31.6	6.8	3496	22	AAH46183
	25	31.6	6.8	6045	22	AAH18244
	26	31.4	6.8	474	22	AAI39048
	27	31.4	6.8	2729	21	AAA38443
	28	31.4	6.8	2730	21	AAZ93780
	29	31.4	6.8	109519	22	AAV08693
	30	31.2	6.7	1044	21	AAZ51697
	31	31.2	6.7	1044	21	AAZ51701
	32	31.2	6.7	1044	21	AAZ51702
	33	31.2	6.7	1044	21	AAZ51703
	34	31.2	6.7	26664	21	AAA60207
	35	31	6.7	905	22	AAH74539
	36	31	6.7	905	22	AAH74541
	37	31	6.7	1294	20	AAZ06824
	38	31	6.7	1294	22	AAH74537
	39	30.8	6.6	1718	18	AAV93529
	40	30.8	6.6	1718	19	AAV18845
	41	30.8	6.6	1718	20	AAV78447
	42	30.8	6.6	1867	18	AAV93530
	43	30.8	6.6	1867	19	AAV18848
	44	30.8	6.6	1867	20	AAV78450
	45	30.8	6.6	106746	21	AAA10225

ALIGNMENTS

RESULT 1	
AAV65766	standard; cDNA; 2614 BP.
ID	AAV65766
XX	AAV65766;
AC	02-FEB-1999 (first entry)
XX	
DT	Rat progression elevated gene-2 (PEG-3) promoter region.
XX	
DE	Progression elevated gene-3; PEG-3; rat; tumour progression;
XX	DNA repair; tumorigenesis; angiogenesis; metastasis; melanoma;
KW	brain cancer; cervix cancer; prostate cancer; lung cancer;
KW	colorectal cancer; neuroblastoma; glioblastoma; diagnosis; therapy;
KW	promoter; ss.
XX	
OS	Rattus sp.
XX	
FH	Key
FT	misc_feature
FT	Location/Qualifiers
FT	128..133
FT	/*tag= a
FT	/label= PEA3
FT	2128..2133
FT	misc_feature
FT	/*tag= b
FT	/label= PEA3
FT	2404..2409
FT	/*tag= c
FT	/label= PEA3
FT	338..344
FT	/*tag= d
FT	/label= E2A
FT	351..357
FT	/*tag= e
FT	misc_feature
FT	/*tag= GRE
FT	/label= GRE

Arabidopsis thalia
EST clone B188. H
Sphingian biosynthe
Sphingomonas genu
Sphingomonas S88 s
Chromosomal fragme
Human low adenosin
Human adenosine re
Human low adenosin
Human low adenosin
Human adenosine re
Mouse Kv1.7 volta
Streptococcus muta
Phosphoryl transpo
Human cDNA sequenc
Probe #7734 used t
Murine desmin gene
Mouse desmin promo
Micromonospora DNA
Burkholderia cepac
Burkholderia multi
Burkholderia multi
Burkholderia multi
Human prostate can
Nucleotide sequenc
Nucleotide sequenc
Streptomyces albid
Nucleotide sequenc
Maturase and Coat
PAR-1 containing b
Bacteriophage ms2
PAR-2 as "Armoured
PAR-2 containing b
HIV gag gene (OSRN
Human PCrTA-1 genom

FT	misc_feature	1094..1099
FT	/tag= f	
FT	/label= GRE	
FT	729..737	
FT	/tag= g	
FT	/label= E2F	
FT	193..197	
FT	/tag= h	
FT	/label= TRE	
FT	489..493	
FT	/tag= i	
FT	/label= TRE	
FT	936..940	
FT	/tag= j	
FT	/label= TRE	
FT	774..779	
FT	/tag= k	
FT	/note= "acute phase reactive regulating element"	
FT	2090..2095	
FT	/tag= l	
FT	/note= "acute phase reactive regulating element"	
FT	2028..2033	
FT	/tag= m	
FT	/label= SP1	
FT	1784..1790	
FT	/tag= n	
FT	/label= AP1	
FT	2552..2558	
FT	/tag= o	
FT	/label= AP1	
FT	865..874	
FT	/tag= p	
FT	/label= AP2	
FT	1597..1609	
FT	/tag= q	
FT	/label= NF1	
XX	WC9842315-A1.	
PN		
XX	01-OCT-1998.	
PD		
XX	20-MAR-1998; 98WO-US05793.	
PF		
XX	21-MAR-1997; 97US-0821818.	
PR		
XX		
PA	(GENO-) GENQUEST INC.	
PA	(UYCO) UNIV COLUMBIA NEW YORK.	
XX		
PI	Fisher PB, Goldstein NI, Su Z, Zhang N;	
PI	WPI; 1998-557025/47.	
DR		
XX		
PS	New isolated Progression Elevated Gene-3 - used to develop products	
XX	for e.g. modulating DNA damage and repair pathways, cancer	
PT	progression or oncogene mediated transformation and angiogenesis.	
PT		
XX	Claim 24; Fig 14A-B; 225pp; English.	
XX		
CC	This is the 5' regulatory region of the rat progression elevated	
CC	gene-3 (PEG-3, see also AAV65764). It was identified using a genomic	
CC	walking strategy, and appears to encompass a functionally complete	
CC	PEG-3 gene promoter including a number of potentially important	
CC	transcriptional motifs. PEG-3 expression is uniquely elevated in	
CC	all cases of rodent progression analysed, and the level of	
CC	expression is also an indicator of DNA damage in that cell. Cells	
CC	in which a reporter gene is under control of the promoter of the	
CC	PEG-3 gene can be used in methods for identifying agents that	
CC	modulate PEF-3 expression or the ability of PEG-3 to induce	
CC	progression, or for determining whether an agent is capable of	
CC	inhibiting DNA damage and repair pathways, cancer progression or	
CC	oncogene-mediated transformation. Compounds that induce DNA	
CC	damage or which regulate angiogenesis can also be identified using	
CC	such cells. The methods can be applied to a progression phenotype	

CC comprising anchorage-independent growth, tumorigenesis, angiogenesis
CC or metastasis, to melanoma, brain, cervical, prostate, lung or
CC colorectal cancer, neuroblastoma or glioblastoma.
XX
SQ Sequence 2614 BP; 737 A; 615 C; 706 G; 556 T; 0 other;

Query Match	92.28;	Score 428;	DB 19;	Length 2614;
Best Local Similarity	99.48;	Pred. No. 1.8e-128;		
Matches 461; Conservative	0;	Mismatches 0;	Indels 3;	Gaps 3;

QY	1	gaaagagaagaatgtggacagcatgttgactgcctgatgaagtltggcgtgtcgtctca	60
Db	1512	gaaagagaagaatgtggacagcatgttgactgcctgatgaagtltggcgtgtcgtctca	1571
QY	61	aagttctgcgagattgcgcgtctcttgatltttagccaagacaacgcctgggaaccag	120
Db	1572	aagttctgcgagattgcgcgtctcttgatltttagccaagacaacgcctgggaaccag	1631
QY	121	gtgacctcacaaaggcccggaatctccgcgagaaattcagtggttllcctctccacc	180
Db	1632	gtgacctcacaaaggcccggaatctccgcgagaaattcagtggttllcctctccacc	1691
QY	181	ttcttcagggacttcggaactccgcctctccggtgacgtcacatalagcgtgcgcaga	240
Db	1692	ttcttcagggacttcggaactccgcctctccggtgacgtcacatalagcgtgcgcaga	1749
QY	241	ctataaactccccgggtgatcgtgtltg'gcgcagatlgactcagtltcgcagcttgttgaaga	300
Db	1750	ctataaactccccgggtgatcgtgtltg'gcgcagatlgactcagtltcgcagcttgttgaaga	1809
QY	301	ttacatgcgagacccccgcgcgactccgcgatcccctltgccgggacagcctltgcgaacgcc	360
Db	1810	ttacatgcgagacccccgcgcgactccgcgatcccctltgccgggacagcctltgcgaacgcc	1869
QY	361	cgtgagacatcacgtcccccgagcccccaacgcctgagggcgacatgaacgcgctgtgccttga	420
Db	1870	cgtgagacatcacgtcccccgagccccca-gcctgagggcgacatgaacgcgctgtgccttga	1928
QY	421	gagcaatccggaaccacgatacgtcttltggcaaacccgaaccggac	464
Db	1929	gagcaatccggaaccacgatacgtcttltggcaaacccgaaccggac	1972

RESULT	2
AAZ23030	
ID	AAZ23030 standard; DNA; 2614 BP.
XX	
AC	AAZ23030;
XX	
DT	17-JAN-2000 (first entry)
XX	
DE	Rat progression elevated gene-3 (PEG-3) promoter sequence.
XX	
KW	Progression-elevated gene-3; PEG-3 gene; regulatory region; cell death;
KW	cell growth; cancer; thymidine kinase; gancyclovir; acyclovir; rat;
KW	promoter; ss.
XX	
OS	Rattus sp.
XX	
PN	WO9949898-A1.
XX	
PD	07-OCT-1999.
XX	
PF	31-MAR-1999; 99WO-US07199.
XX	
PR	31-MAR-1998; 98US-0052753.
XX	
PA	(UYCO) UNIV COLUMBIA NEW YORK.
XX	
PI	Fisher PB;
XX	
DR	WPI; 1999-591184/50.

XX Novel vectors useful for studying the progression of cancer -
PT
XX
PS Disclosure; Fig 14 A-B; 251pp; English.
XX
CC The invention relates to an inducible progression-elevated gene-3 (PEG-3
CC gene) regulatory region functionally linked to a gene encoding a product
CC that causes or may be induced to cause the death or inhibition of cancer
CC cell growth. A vector of the invention which contains a gene encoding
CC thymidine kinase or a product which causes the cell to express a
CC specific antigen can be administered along with gancyclovir or acyclovir,
CC or an antibody or fragment to the antigen, respectively, to treat cancer
CC in a subject. The PEG-3 gene is useful for generating new cloning and
CC expression vectors, transfected cells, and for developing methods for
CC cultured growth of such cells. The PEG-3 polynucleotide is also useful as
CC a source of primers and probes to study the progression of cancer, and to
CC detect the presence of the gene. The present sequence represents the
XX nucleotide sequence of the rat PEG-3 gene promoter.
SQ Sequence 2614 BP; 737 A; 615 C; 706 G; 556 T; 0 other;

Query Match	92.28;	Score 428;	DB 20;	Length 2614;
Best Local Similarity	99.48;	Pred. No. 1.8e-128;		
Matches 461; Conservative	0;	Mismatches 0;	Indels 3;	Gaps 3;

QY	1	gaaagagaagaatctgggaacagcatgttacttgccctgatgaagtggcgtgtctgtctcaa	60
Db	1512	gaaagagaagaatctgggaacagcatgttacttgccctgatgaagtggcgtgtctgtctcaa	1571
QY	61	aagttctgcgagatttcagcgctctcttggaatttgaccaaaggacaacgccttgggaagccaag	120
Db	1572	aagttctgcgagatttcagcgctctcttggaatttgaccaaaggacaacgccttgggaagccaag	1631
QY	121	gtgacctcacaaagccccggaaatctcgcgcgagaatttcagtgtytttccctctccaac	180
Db	1632	gtgacctcacaaagccccggaaatctcgcgcgagaatttcagtgtytttccctctccaac	1691
QY	181	ttcttcagggacttcggaactccgcctctccgcgtgacqtlcagcatagacgtgcgtcaaga	240
Db	1692	ttcttcagggacttcggaactccgcctctccgcgtgacqtlcagcatagacgtgcgtcaaga	1749
QY	241	ctataaactcccgggtgatcgtgtgtggcgcagattgacccagttcgcagcttgttgyaaga	300
Db	1750	ctataaactcccgggtgatcgtgtgtggcgcagattgacccagttcgcagcttgttgyaaga	1809
QY	301	ttacatgcgagacccccgcgcgactccgcatalccctttgccggygacagcctttgcgacagcc	360
Db	1810	ttacatgcgagacccccgcgcgactccgcatalccctttgccggygacagcctttgcgacagcc	1869
QY	361	cgtgagacatcacgctcccccgagcccccaacgccttgagggcgacatgaacgcgctggccttga	420
Db	1870	cgtgagacatcacgctcccccgagccccca-gcctgagggcgacatgaacgcgctggccttga	1928
QY	421	gagcaatccggacccaacgatcgcttttggcaaaaccggaaccggac	464
Db	1929	gagcaatccggacccaacgatcgcttttggcaaaaccggaaccggac	1972

RESULT	3
AAV65765	
ID	AAV65765 standard; cDNA; 2111 BP.
XX	
AC	AAV65765;
XX	
DT	02-FEB-1999 (first entry)
XX	
DE	Human progression elevated gene-2 (PEG-3) cDNA.
XX	
KW	Progression elevated gene-3; PEG-3; human; tumour progression;
KW	DNA repair; tumorigenesis; angiogenesis; metastasis; melanoma;
KW	brain cancer; cervix cancer; prostate cancer; lung cancer;
KW	colorectal cancer; neuroblastoma; glioblastoma; diagnosis; therapy;

KW	SS.	
XX		
OS	Homo sapiens.	
XX		
PH	Key	Location/Qualifiers
FT	CDS	294..2030
FT		/*tag= a

PN	WO9842315-A1.
XX	
PD	01-OCT-1998.
XX	
PF	20-MAR-1998; 98WO-US05793.
XX	
PR	21-MAR-1997; 97US-0821818.
XX	
PA	(GENO-) GENQUEST INC.
PA	(UYCO) UNIV COLUMBIA NEW YORK.
XX	
PI	Fisher PB, Goldstein NI, Su Z, Zhang N;
XX	
DR	WPI; 1998-557025/47.
DR	P-PSDB; AAW79958.

New Isolated Progression Elevated Gene-3 - used to develop products for e.g. modulating DNA damage and repair pathways, cancer progression or oncogene mediated transformation and angiogenesis.

Claim 4; Fig 13A-C; 225pp; English.

AA This nucleotide sequence includes an open reading frame encoding
CC human progression elevated gene-3 (PEG-3) protein (see AAW79957).
CC PEG-3 clones were isolated from an MCF-7 cDNA library using rat
CC PEG-3 cDNA (see AAV65764) as probe. 5'RACE (see AAV65767-68) was used
CC to generate full-length cDNA. PEG-3 mRNA is highly expressed in
CC most human tumour cell lines. Probes derived from the gene can be
CC used to monitor tumour progression. Antisense sequences can be
CC used to inhibit tumour progression. Cells in which a reporter gene
CC is under control of the promoter (see AAV65766) of the rat PEG-3
CC can be used in methods for identifying agents that modulate PEF-3
CC expression or the ability of PEG-3 to induce progression, or for
CC determining whether an agent is capable of inhibiting DNA damage
CC and repair pathways, cancer progression or oncogene-mediated
CC transformation. Compounds that induce DNA damage or which regulate
CC angiogenesis can also be identified using such cells. Transgenic
CC animals and vaccines comprising PEG-3 polypeptides and an immune
CC response enhancer are also claimed. The methods can be applied to
CC a progression phenotype comprising anchorage-independent growth,
CC tumorigenesis, angiogenesis or metastasis, to melanoma, brain,
CC cervical, prostate, lung or colorectal cancer, neuroblastoma or
CC glioblastoma.

Query Match	28.3%;	Score 131.4;	DB 19;	Length 211;
Best Local Similarity	99.2%;	Pred. No. 1.8e-32;		
Matches 132; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

[illegible]

RESULT 4

AAZ23029
ID AAZ23029 standard; cDNA; 2111 BP.
XX
AC AAZ23029;
XX
DT 17-JAN-2000 (first entry)
XX
DE Human progression elevated gene-3 (PEG-3) cDNA sequence.
XX
KW Progression-elevated gene-3; PEG-3 gene; regulatory region; cell death;
KW cell growth; cancer; thymidine kinase; gancyclovir; acyclovir; human; ss.
XX
OS Homo sapiens.
XX
PN WO9949898-A1.
XX
PD 07-OCT-1999.
XX
PF 31-MAR-1999; 99WO-US07199.
XX
PR 31-MAR-1998; 98US-0052753.
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI Fisher PB;
XX
DR WPI; 1999-591184/50.
DR P-PSDB; AAY41104.
XX
PT Novel vectors useful for studying the progression of cancer -
XX
PS Disclosure; Fig 13A-C; 251pp; English.
XX
CC The invention relates to an inducible progression-elevated gene-3 (PEG-3
CC gene) regulatory region functionally linked to a gene encoding a product
CC that causes or may be induced to cause the death or inhibition of cancer
CC cell growth. A vector of the invention which contains a gene encoding
CC thymidine kinase or a product which causes the cell to express a
CC specific antigen can be administered along with gancyclovir or acyclovir,
CC or an antibody or fragment to the antigen, respectively, to treat cancer
CC in a subject. The PEG-3 gene is useful for generating new cloning and
CC expression vectors, transfected cells, and for developing methods for
CC cultured growth of such cells. The PEG-3 polynucleotide is also useful as
CC a source of primers and probes to study the progression of cancer, and to
CC detect the presence of the gene. The present sequence represents the
CC nucleotide sequence of the human PEG-3 cDNA.
XX
SQ Sequence 2111 BP; 533 A; 608 C; 575 G; 395 T; 0 other;

Query Match 28.3%; Score 131.4; DB 20; Length 2111;
Best Local Similarity 99.2%; Pred. No. 1.8e-32;
Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 332 ccttgccggagacagcctttgcgacagcccggtgagacatcacgtcccgagccccaagcc 391
DB 120 ccttgccggagacagcctttgcgacagcccggtgagacatcacgtcccgagccccaagcc 179
QY 392 tgaaggcgacatgaacgcgctggtccttgagagcaatccgagcccaagatcgctttggca 451
DB 180 tgaaggcgacatgaacgcgctggtccttgagagcaatccgagcccaagatcgctttggca 239
QY 452 aacggaaccggac 464
DB 240 aacggaaccggac 252

RESULT 5
ID AAX87383 standard; cDNA; 2111 BP.
XX
AC AAX87383;
XX

DT 08-OCT-1999 (first entry)
XX
DE Human progression elevated gene-3 (PEG-3) cDNA.
XX
KW Progression elevated gene-3; PEG-3; human; angiogenesis; cancer;
KW therapy; vaccine; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 294..2030
FT /*tag= a
XX
PN WO9937776-A1.
XX
PD 29-JUL-1999.
XX
PF 26-JAN-1999; 99WO-US01623.
XX
PR 29-JAN-1998; 98US-0072941.
PR 26-JAN-1998; 98US-0072469.
XX
PA (GENO-) GENQUEST INC.
XX
PI Fisher PB, Zhang N;
XX
DR WPI; 1999-458694/38.
DR P-PSDB; AAY06514.
XX
PT Modulation of angiogenesis by altering the expression and/or
PT activity of a progression-associated protein, especially for cancer
PT treatment
XX
PS Claim 1; Page 71-74; 81pp; English.
XX
CC This is the nucleotide sequence of a cDNA clone corresponding to
CC the human progression elevated gene-3 (PEG-3). It codes for a
CC 578-amino acid protein (see AAY06514), and represents a new member
CC of the gadd34/MyD16 gene family. The human PEG-3 cDNA was
CC isolated from a MCF-7 cDNA library using rat PEG-3 (see AAX87387) as
CC probe. Human PEG-3 is generally expressed in cells that are in
CC progression, including most human tumour cell lines. A claimed
CC method for modulating angiogenesis in an organism comprises
CC administering an agent that alters expression and/or activity of a
CC PEG-3 protein, such as an antisense polynucleotide or antibody.
CC Also claimed is a method for determining whether an agent
CC modulates angiogenesis, where the candidate agent is present
CC within a combinatorial small molecule library. Vaccines and
CC pharmaceutical compositions comprising such compounds are also
CC provided and may be used to prevent angiogenesis, especially
CC related to cancer cell progression.
XX
SQ Sequence 2111 BP; 533 A; 608 C; 575 G; 395 T; 0 other;

Query Match 28.3%; Score 131.4; DB 20; Length 2111;
Best Local Similarity 99.2%; Pred. No. 1.8e-32;
Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 332 ccttgccggagacagcctttgcgacagcccggtgagacatcacgtcccgagccccaagcc 391
DB 120 ccttgccggagacagcctttgcgacagcccggtgagacatcacgtcccgagccccaagcc 179
QY 392 tgaaggcgacatgaacgcgctggtccttgagagcaatccgagcccaagatcgctttggca 451
DB 180 tgaaggcgacatgaacgcgctggtccttgagagcaatccgagcccaagatcgctttggca 239
QY 452 aacggaaccggac 464
DB 240 aacggaaccggac 252

RESULT 6

AAC09789	ID	AAC09789 standard; cDNA; 287 BP.
XX	AC	AAC09789;
XX	DT	06-OCT-2000 (first entry)
XX	DE	Human secreted protein 5' EST, SEQ ID NO: 13864.
XX	KW	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW	gene therapy; chromosome mapping; ss.	
XX	OS	Homo sapiens.
XX	PN	EP1033401-A2.
PD	XX	06-SEP-2000.
XX	PF	21-FEB-2000; 2000EP-0200610.
XX	PR	26-FEB-1999; 99US-0122487.
PA	(GEST) GENSET.	
PI	Dumas Milne Edwards J, Duclert A, Giordano J;	
DR	WP1; 2000-500381/45.	
XX	New nucleic acid that is a 5' expressed sequence tag (5' EST) for	
PT	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for	
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures -	
PS	Claim 1; SEQ ID 13864; 71pp + CD-ROM; English.	
CC	The present sequence is one of a large number of 5' ESTs derived from	
CC	mRNAs encoding secreted proteins. No ORF has yet been conclusively	
CC	identified within the present sequence. The 5' ESTs were prepared from	
CC	total human RNAs or polyA+ RNAs derived from 30 different tissues. EST	
CC	sequences usually correspond mainly to the 3' untranslated region (UTR)	
CC	of the mRNA because they are often obtained from oligo-dT primed cDNA	
CC	libraries. Such ESTs are not well suited for isolating cDNA sequences	
CC	derived from the 5' ends of mRNAs and even in those cases where longer	
CC	cDNA sequences have been obtained, the full 5' UTR is rarely included.	
CC	5' ESTs are derived from mRNAs with intact 5' ends and can therefore be	
CC	used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used	
CC	in diagnostic, forensic, gene therapy and chromosome mapping procedures.	
CC	They are used to obtain upstream regulatory sequences and to design	
CC	expression and secretion vectors.	
XX	Sequence 287 BP; 52 A; 101 C; 79 G; 54 T; 1 other:	
SQ		
	Query Match	17.6%; Score 81.8; DB 21; Length 287;
	Best Local Similarity	66.4%; Pred. No. 9e-17;
	Matches 148; Conservative	0; Mismatches 72; Indels 3; Gaps 2;
OY	243 ataactccggytgatcgtgtggcgagatgaactcagtgcga-gcttgygaagat 301	
DB	1 ataaagcctagtggccattgttcgtcgtctctatcgtttcccatccagltgtgat 60	
OY	302 tacatgcgagacccccgcgcgactccgcatacccttgcgcggagacagccttgcagacgcc 361	
DB	61 cttaatgaagaacgctgcagcacaccgcgccgcttgtgcacagcacttgaggcagccg 120	
OY	362 gttagacatcaagt--cccagagccccaagccttgaggcgacatgaacgcgtgaccttg 419	
DB	121 gagatactctgagttaactcggaagccgaagcctgaggtgatgaatgaacgcgtgacctcc 180	
OY	420 agagcaatccgcgaccacagatcgcttcttggaacaacggaacgg 462	
DB	181 ctaacgcgtccggaacctgtgatcgcttcttggcagacccaacgg 223	

RESULT	7
ID	AAH05790 standard; cDNA; 642 BP.
XX	AAH05790;
XX	26-JUN-2001 (first entry)
DT	
DE	Human cDNA clone (5'-primer) SEQ ID NO:2625.
XX	
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
OS	Homo sapiens.
XX	
PN	EP1074617-A2.
XX	
PD	07-FEB-2001.
XX	
PF	28-JUL-2000; 2000EP-0116126.
XX	
PR	29-JUL-1999; 99JP-0248036.
PR	27-AUG-1999; 99JP-0300253.
PR	11-JAN-2000; 2000JP-0118776.
PR	02-MAY-2000; 2000JP-0183767.
PR	09-JUN-2000; 2000JP-0241899.
PA	(HELI-) HELIX RES INST.
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR	WPI; 2001-318749/34.
XX	
PT	Primer sets for synthesizing polynucleotides, particularly the 5602
PT	full-length cDNAs defined in the specification, and for the detection
PT	and/or diagnosis of the abnormality of the proteins encoded by the
PT	full-length cDNAs -
XX	
PS	Claim 1; SEQ ID 2625; 2537bp + CD ROM; English.
XX	
CC	The present invention describes primer sets for synthesising 5602
CC	full-length cDNAs defined in the specification. Where a primer set
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC	to the complementary strand of a polynucleotide which comprises one of
CC	the 5602 nucleotide sequences defined in the specification, where the
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC	of an oligonucleotide comprising a sequence complementary to the
CC	complementary strand of a polynucleotide which comprises a 5'-end
CC	sequence and an oligonucleotide comprising a sequence complementary to a
CC	polynucleotide which comprises a 3'-end sequence, where the
CC	oligonucleotide comprises at least 15 nucleotides and the combination of
CC	the 5'-end sequence/3'-end sequence is selected from those defined in
CC	the specification. The primer sets can be used in antisense therapy and
CC	in gene therapy. The primers are useful for synthesising polynucleotides,
CC	particularly full-length cDNAs. The primers are also useful for the
CC	detection and/or diagnosis of the abnormality of the proteins encoded by
CC	the full-length cDNAs. The primers allow obtaining of the full-length
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC	AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC	represent oligonucleotides, all of which are used in the exemplification
CC	of the present invention.
XX	
SQ	Sequence 642 BP; 131 A; 200 C; 190 G; 115 T; 6 other;
Query Match	17.5%; Score 81.2; DB 22; Length 642;
Best Local Similarity	70.9%; Pred. No. 2e-16;
Matches 122; Conservative	0; Mismatches 48; Indels 2; Gaps 1;
Oy	293 gtgaagattacatgcgagaccgccggcgcatccgttgcggagacaccttg 352 23 gttgttatcttatgcagaacgctgcacgacccccgcgcgcttgtcgcacgycacttg 82

QY 353 cgacagcccgctgtgagacatcaagt--ccccgaagcccacaagcgtgtgaggcgacatgaacgcg 410
| | | | | | | | | | | | | | | | | | | | | |
Db 83 agcgacggcgagataactcttgatgttactcggagcccgacgccttgaggtgtgatgaacycg 142

QY 411 ctggtccttgaggaagaatccgcgaccacaagatgccttttgcaaacgcgaaccgg 462
| | | | | | | | | | | | | | | | | | | | | |
Db 143 ctgtgccttcctaaccgctccgcgacctgtgatcgccttctgtgcgaccgcgaaccgg 194

RESULT	8
AAZ99766	
ID AAZ99766	standard; cDNA; 2331 BP.

AC AA299766;

DT 12-JUL-2000 (first entry)

DE cDNA encoding a human apoptosis associated protein (GADD34).

KW Human; apoptosis associated protein; GADD34; antiviral agent;
 KW growth arrest and DNA damage-inducible gene 34; viral infection;
 KW serine/threonine phosphatase; ss.

OS Homo sapiens.

FH	Key	Location/Qualifiers
FT	CDS	223..2247

FT /product= "apoptosis associated protein"

PN GB2342716-A.

PD 19-APR-2000.

PF 27-JUL-1999; 99GB-0017631.

PR 14.-SEP-1998; 98GB-0020025.

PA (PEIZ) PEIZER LTD.

PI Ciaramella G;

DR WPI; 2000-285397/25.
DR P-PSDB; AAY84366.

PT Screening assay for potential antiviral agents that modify
PT serine/threonine phosphatase activity for treating human viral
PT infections -

PS Claim 7; Page 98; 106pp; English.

The present sequence encodes a human apoptosis associated protein (GADD34). It is encoded by a growth arrest and DNA damage-inducible gene 34. It is used in an assay method for identifying an anti-viral agent that can affect the activity or expression of a nucleotide sequence or its expression product. The assay method comprises contacting an agent with a nucleotide sequence coding for a serine/threonine phosphatase or its expression product, and determining whether the agent affects the activity or expression of the nucleotide sequence. The assay is used for screening for antiviral agents useful in the treatment of human viral infections. The anti-viral modulators of type 1 protein phosphatase activity inhibit the action of a host component which interferes with the activation of a host-cell interferon-induced, double-stranded RNA-activated protein kinase e.g. p68, eIF2 or eIF2alpha.

Sequence 2331 BP; 543 A; 666 C; 712 G; 410 T; 0 other;

Query Match	17.5%;	Score 81.2;	DB 21;	length 2331;
Best Local Similarity	70.9%;	Pred. No. 3.4e-16;		
Matches 122; Conservative	0;	Mismatches 48;	Indels 2;	Gaps 1;

```

Qy      293  gtgtaagattacatgtcgagacccccgcgcactccgcataccctttgccggagacagccctttg 352
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      5    gtgtgtgatcttatgtcaagacgctgcacgacccccgcgccttgcgcacagcgacattg 64

```

```
QY      353 cgacagcccgctgagacatcacgt--ccccagccccacgcctgaaggcagacatgaacgcg 410
        | ||||| | || || | | ||||| ||||||||| || |||||||||
Db      65 aggcagccggagataactcttgatttacttcgagcccgacgccctgaagggtgatgatgaacgcg 124
```

```
Oy 411 ctggcccttgagcaatccgcgacccacgatacgccttltgccaaccgaaaccg 462
      ||||| | | ||||| ||||| ||||| ||||| |||||
Db 125 ctggcctccctaaccgctccgagactgtgatcgcttcttgccagaccgaaaccg 176
```

RESULT 9
AAX19798

AC AAX19798;

DT 09-JUN-1999 (first entry)

DE Human growth regulator protein GRREG encoding cDNA.

KW Human; growth regulator protein; GRREG; cancer; ss.

OS Homo sapiens.

EH	Key	Location/Qualifiers
FT	CDS	226..2250

PN W09902680-A1.

PD 21-JAN-1999.

PF 30-JUN-1998; 98WO-US13409.

PR 11-JUL-1997; 97US-0893852.

PA (INCY-) INCYTE PHARM INC.

PI Bandman O, Corley NC, Lal P, Shah P;

DR WPI: 1999-120881/10.

PT Novel human growth regulator protein - useful in the treatment of
PT cancers

PS Claim 7; Fig 1; 70pp; English.

CC The present sequence encodes a human growth regulator protein,
CC designated GRREG. The protein can be used in methods to treat cancer.
CC The protein can also be used in binding assays to detect (ant)agonists
CC of GRREG. A fragment from the nucleic acid sequence encoding GRREG can
CC be used as a probe for detecting GRREG encoding sequences (especially
CC in PCR amplified samples).

Sequence 2344 BP; 556 A; 667 C; 709 G; 412 T; 0 other;

Query Match	17.5%;	Score 81.2;	DB 20;	Length 2344;
Best Local Similarity	70.9%;	Pred. No. 3.4e-16;		
Matches 122; Conservative	0;	Mismatches 48;	Indels 2;	Gaps 1;

Qy 293 gtggaagatlaacatgcgagacccccgcgcgcgactccgcacatccctttgcgcggagacagcccttg 352
||| ||| |||| |||| | |||| |||| || |||| || |||| ||||
Db 8 gttgttgcattctatgcgagacgcgtgcacgacccccgcgcgcgcgtgtctgcacagcgacattg 67

Qy 353 cgacagcccgtagacatacaagt--ccccgagcccacgccttaaggcgaatgaacgcg 410
| | | | | | | | | | | | | | | | | | | | | |
Db 68 aggcagccggagaataactctgaatttacctgagcccgacgcctgagsgtgagatgaacgcg 127

PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match 8.0%; Score 37; DB 21; Length 1606;
Best Local Similarity 57.3%; Pred. No. 0.059;
Matches 67; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY	37	gatgaattggcgtgtctgtctcaaaagttctgcgagattgacgctctctgattgagc	96
Db	267	GAAGACGTTGGCGATCGTGTTCACAGACGTCACAGATTGAGTTCTTGTACGTGTCAAA	208
QY	97	caaggaacagccttgggaagccacggtgacctcacaaagcccgaaatctccgagaa	153
Db	207	AAAGTACAGGCGCTTGCAAGCCCATTTAGCTTCATCGTTCCGGAATTGAGAGAGAA	151

RESULT 13
AAV86970
ID AAV86970 standard; cDNA; 494 BP.
XX
AC AAV86970;
DT 27-APR-1999 (first entry)
DE EST clone B188.
XX
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokines; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS Homo sapiens.
XX
PN WO9845435-A2.
XX 15-OCT-1998.
PD
PF 10-APR-1998; 98WO-US06954.
XX
PR 10-APR-1997; 97US-0835913.
XX
PA (GENE) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX WPI; 1999-070076/06.
DR
XX
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
XX
PS Claim 1; Page 409; 633pp; English.
XX
CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
XX
SQ Sequence 494 BP; 158 A; 88 C; 91 G; 157 T; 0 other;

QY	20	acagcatgtactgcctgatgaaglttggcgtgtgtcttcaaaagttctgcgagattgacg	79
Db	100	agattataagacatcccttaagagtttaaatgtctctgaacaagcataacagcgtctagt	159

Query Match 7.8%; Score 36.2; DB 20; Length 494;
Best Local Similarity 48.3%; Pred. No. 0.066;
Matches 101; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

CC sphingan carbohydrates. DNA fragments of S88 can be inserted into
CC a vector in multiple copies and used to produce engineered bacteria
CC that are hyper-producers of sphingan.
XX
SQ Sequence 28804 BP; 4974 A; 9806 C; 9228 G; 4796 T; 0 other;

Query Match 7.3%; Score 33.8; DB 17; Length 28804;
Best Local Similarity 51.0%; Pred. No. 2.1;
Matches 80; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 301 ttacatgagagaccccgcgactccgcatccctttgcccggagacagcctttgcgacagcc 360
||| ||| || | | | | | | | | | | | | | | | |
Db 24702 ttaccgcgcgccgcttcacgtcgtcgcagtggtacgccacctcgttcgccaggcg 24761
QY 361 cgtgagacatcacgtcccgagcccaagcctgaggcgacatgaacgctggccttga 420
|| | | | | | | | | | | | | | | | | | | | |
Db 24762 cgcgcgtcatcatctcgacaagctcacctatgcggcaacccgctcgtcgtgacccga 24821
Y 421 gagcaatccggaccacgactcgttttggcaaccga 457
| | | | | | | | | | | | | | | | | | | | |
Db 24822 tcgagaacgcccccaactaccgcttcgtccaagccga 24858

RESULT 15

AAT92474
ID AAT92474 standard; DNA; 28804 BP.

AC AAT92474;

DT 04-FEB-1998 (first entry)

DE Sphingomonas genus microbe isolated DNA sequence producing sphingan.

KW Sphingomonas microbe; sphingan polysaccharide biosynthesis gene;
KW sphingan S-88; spsB gene; ss.

OS Sphingomonas sp.

PN JP09252775-A.

PD 30-SEP-1997.

PF 24-JAN-1996; 96JP-0043977.

PR 16-JAN-1996; 96JP-0004621.

PR 24-JAN-1995; 95US-0377440.

PA (SHIN-) SHINETSU BIO INC.

PA (SHIE) SHINETSU CHEM IND CO LTD.

DR WPI; 1997-530148/49.

PT Sphingan polysaccharide biosynthesis gene from Sphingomonas microbe
PT - can be introduced into Sphingomonas host cells to increase
PT sphingan production

PS Claim 74; Page 34-43; 55pp; Japanese.

XX The present sequence represents a new DNA sequence which has been
CC isolated from a sphingan polysaccharide-producing microbe of the
CC genus Sphingomonas. When recombined in a Sphingomonas host, the
CC presence of several copies of the sphingan polysaccharide biosynthesis
CC gene results in a microbe which expresses high levels of sphingan.
CC A method has been developed for producing a Sphingomonas microbe with
CC enhanced sphingan polysaccharide expression, by transforming the cell
CC with the sphingan polysaccharide biosynthesis gene. The method can
CC produce sphingan in large quantities.

SQ Sequence 28804 BP; 4974 A; 9804 C; 9230 G; 4796 T; 0 other;

Query Match 7.3%; Score 33.8; DB 18; Length 28804;

Best Local Similarity 51.0%; Pred. No. 2.1;
Matches 80; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 301 ttacatgagagaccccgcgactccgcatccctttgcccggagacagcctttgcgacagcc 360
||| ||| || | | | | | | | | | | | | | | | |
Db 24702 ttaccgcgcgccgcttcacgtcgtcgcagtggtacgccacctcgttcgccaggcg 24761
QY 361 cgtgagacatcacgtcccgagcccaagcctgaggcgacatgaacgctggccttga 420
|| | | | | | | | | | | | | | | | | | | | |
Db 24762 cgcgcgtcatcatctcgacaagctcacctatgcggcaacccgctcgtcgtgacccga 24821
QY 421 gagcaatccggaccacgactcgttttggcaaccga 457
| | | | | | | | | | | | | | | | | | | | |
Db 24822 tcgagaacgcccccaactaccgcttcgtccaagccga 24858

Search completed: April 5, 2002, 08:51:24
Job time: 1707 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2002, 08:22:57 ; Search time 1472.67 Seconds
(without alignments)
5197.832 Million cell updates/sec

Title: US-09-621-781-1_COPY_1507_1970
Perfect score: 464
Sequence: 1 gaaagagaagagaatygga.....tttgcaaacggaaccggac 464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	464	100.0	1940	10	AF351130	AF351130 Rattus no
2	300.6	64.8	215734	2	AC073828	AC073828 Mus muscu
3	154	33.2	216	10	MMU83984	U83984 Mus musculu
4	136.6	29.4	2275	10	MMMDPRMR	X51829 Mouse myelo
5	116.4	25.1	179538	2	AC024740	AC024740 Homo sapi
6	114.8	24.7	110000	2	AC008749_3	Continuation (4 of
7	114.8	24.7	231450	2	AC026803	AC026803 Homo sapi
8	114.2	24.6	2088	10	CRUGADD34X	L28147 Hamster Gad
9	106.2	22.9	165166	2	AC064877	AC064877 Homo sapi
10	96.2	20.7	224	10	CGU83983	U83983 Cricetulus
11	81.2	17.5	2331	9	HSU83981	U83981 Homo sapien
12	81.2	17.5	2344	6	ARI00139	ARI00139 Sequence
13	81.2	17.5	2349	9	AK001361	AK001361 Homo sapi
14	81.2	17.5	2358	9	BC003067	BC003067 Homo sapi
15	45.4	9.8	233	9	HSU83982	U83982 Homo sapien
16	37.8	8.1	151498	9	AC027139	AC027139 Homo sapi
17	37.4	8.1	202475	2	AC009973	AC009973 Homo sapi
18	37	8.0	91566	8	F25C20	AF297204 Bos tauru
19	36.6	7.9	2397	9	HSM801096	AF326988 Cithidia
20	36.6	7.9	3993	9	HSM802091	AC027810 Homo sapi
21	36.6	7.9	4241	9	AB020662	AC024619 Homo sapi
22	36.6	7.9	4578	9	AF204231	AL137686 Homo sapi
23	36.6	7.9	174320	2	AC021822	AB020662 Homo sapi
24	36.6	7.9	192361	2	AC025678	AF204231 Homo sapi
25	36.2	7.8	4371	9	AF164622	AC021822 Homo sapi
26	35.4	7.6	617	11	AF297204	AF164622 Homo sapi
27	35.2	7.6	1218	3	AF326988	AF297204 Bos tauru
28	35	7.5	155628	2	AC027810	AF326988 Cithidia
29	35	7.5	203397	2	AC024619	AC027810 Homo sapi
30	34.8	7.5	148098	2	AL359315	AC024619 Homo sapi
31	34.8	7.5	152623	9	AC011595	AL359315 Homo sapi
32	34.8	7.5	164388	2	AL591647	AC011595 Homo sapi
33	34.8	7.5	173829	2	AL451077	AL591647 Homo sapi
34	34.8	7.5	181427	2	AC027209	AL451077 Homo sapi
35	34.8	7.5	183763	2	AC026900	AC027209 Homo sapi
36	34.8	7.5	192458	2	AL592284	AC026900 Homo sapi
37	34.8	7.5	219461	2	AC087864	AL592284 Homo sapi
38	34.8	7.5	227243	9	AC008569	AC087864 Homo sapi
39	34.6	7.5	6426	1	RCBLUGNS	AC008569 Homo sapi
40	34.6	7.5	113606	8	AC087192	Z46611 R. capsulatu
41	34.6	7.5	189370	1	AF010496	AC087192 Oryza sat
42	34.4	7.4	15019	1	AE006957	AF010496 Rhodobact
43	34.4	7.4	29550	1	MTV039	AE006957 Mycobacte
44	34.4	7.4	124219	9	HSJ196B13	AL021942 Mycobacte
45	34.4	7.4	147165	2	AC023179	AL121968 Human DNA

ALIGNMENTS

RESULT 1
AF351130 1940 bp DNA 09-APR-2001
LOCUS Rattus norvegicus progression elevated gene-3 protein (Peg-3) gene,
DEFINITION promoter region.
ACCESSION AF351130
VERSION AF351130
KEYWORDS AF351130.1 GI:13562027
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 1940)
AUTHORS Su,Z., Shi,Y. and Fisher,P.B.
TITLE Cooperation between API and PEA3 sites within the progression
elevated gene-3 (PEG-3) promoter regulate basal and differential
expression of PEG-3 during progression of the oncogenic phenotype
in transformed rat embryo cells
JOURNAL Oncogene 19 (30), 3411-3421 (2000)

MEDLINE 20374129
PUBMED 10918598
REFERENCE 2 (bases 1 to 1940)
AUTHORS Su,Z., Shi,Y. and Fisher,P.B.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2001) Urology, Columbia University, 630 West 168th Street, New York, NY 10032, USA

FEATURES
source
1. 1940
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
1. 1940
/gene="Peg-3"
1. >1940
/gene="Peg-3"
/note="progression elevated gene-3"

BASE COUNT 568 a 496 c 468 g 408 t
ORIGIN

Query Match 100.0%; Score 464; DB 10; Length 1940;
Best Local Similarity 100.0%; Pred. No. 1.7e-117;
Matches 464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaagagaaagaatggtgacagcatgtgactgcctgatgaattggcgtgtgtctcaa 60
Db 1477 GAAAGAGAAAGAAATGGGACAGCATGTGACTGCCCTGATGAATTGGCGTGTGCTCAA 1536

QY 61 aagttctgcgaattgaagcgtctctgtgatttgagccaagacacgcctgggaagccacg 120
Db 1537 AAGTCTCGCAGATTGACGGCTCTCTGTGATTGAGCCAAGGACACGCTGGGAAGCCACG 1596

QY 121 gtgacctcaagaagcccggaatcttcgcgcagaatttcagtgtttctctctcacc 180
Db 1597 GTGACCTCAAGGCGCCGGAATCTCCGAGAAATTTCAGTGTGTTTCTCTCCACACC 1656

QY 181 ttctcaggaacttcgaaactccgcctctccggtgacgtcagcatagcgtgcgtcaga 240
Db 1657 TTTCTCAGGGAAGCTCCGAAACTCCGCTCTCCGGTGACGTCAGCATAGCGCTGCTCAGA 1716

QY 241 ctataaactcccggtgtgacgtgtgtggcgagattgactcagttcgcagcttgtgaaga 300
Db 1717 CTATAAAGTCCCGGTGATCGTGTGGCGCAGATTGACTCAGTTCCGAGCTTGTGAAGA 1776

QY 301 ttacatgcgagaccccgcgactccgcattcccttgcgcggagacagccttgcgacagcc 360
Db 1777 TTACATGCGAGACCCCGCGGACTCCGCATCCCTTTGCCGGGACAGCCTTTGCCAGAGCC 1836

QY 361 cgtgagacatcacgtcccccagaccccaagcctgagggcgacatgaacgcgtgccttga 420
Db 1837 CGTGAGACATCACGTCCCGGAGCCCGCAGCCTGAGGGGACATGAACGCGCTGCGCTTGA 1896

QY 421 gagcaatccggacccacgactcttggcaaacggaacccgac 464
Db 1897 GAGCAATCCGGACCCACGATCGCTTTGGCAACCGAACCAGCCGGAC 1940

RESULT 2
AC073828 215734 bp DNA HTG 29-JUN-2000
LOCUS Mus musculus clone RP23-9J18, WORKING DRAFT SEQUENCE, 21 unordered
DEFINITION pieces.
AC073828
AC073828.1 GI:8810445
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 215734)
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Mouse
Unpublished

REFERENCE 2 (bases 1 to 215734)
DOE Joint Genome Institute.
AUTHORS Direct Submission
TITLE Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT
-----Genome Center
Center: Joint Genome Institute
Center code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 1740562
Center clone name: RPCI-23_9J18

Summary Statistics
Consensus quality: 196667 bases at least Q40
Consensus quality: 207971 bases at least Q30
Consensus quality: 209928 bases at least Q20
Estimated insert size: 207740; agarose-fp estimation
Estimated insert size: 213734; sum-of-contigs estimation
Quality coverage: 6.12 in Q20 bases; agarose-fp estimation
Quality coverage: 5.94 in Q20 bases; sum-of-contigs estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1013: contig of 1013 bp in length
1014 1113: gap of unknown length
1114 2175: contig of 1062 bp in length
2176 2275: gap of unknown length
2276 3429: contig of 1154 bp in length
3430 3529: gap of unknown length
3530 4884: contig of 1355 bp in length
4885 4984: gap of unknown length
4985 6848: contig of 1864 bp in length
6849 6948: gap of unknown length
6949 9317: contig of 2369 bp in length
9318 9417: gap of unknown length
9418 12128: contig of 2711 bp in length
12129 12228: gap of unknown length
12229 14970: contig of 2742 bp in length
14971 15070: gap of unknown length
15071 19571: contig of 4501 bp in length
19572 19671: gap of unknown length
19672 25961: contig of 6290 bp in length
25962 26061: gap of unknown length
26062 36753: contig of 10692 bp in length
36754 36853: gap of unknown length
36854 49116: contig of 12263 bp in length
49117 49216: gap of unknown length
49217 61879: contig of 12663 bp in length
61880 61979: gap of unknown length
61980 72513: contig of 10534 bp in length
72514 72613: gap of unknown length
72614 83721: contig of 11108 bp in length
83722 83821: gap of unknown length
83822 98633: contig of 14812 bp in length
98634 98733: gap of unknown length
98734 118323: contig of 19590 bp in length
118324 118423: gap of unknown length
118424 137668: contig of 19245 bp in length
137669 137768: gap of unknown length
137769 162338: contig of 24570 bp in length
162339 162438: gap of unknown length
162439 186625: contig of 24187 bp in length
186626 186725: gap of unknown length
186726 215734: contig of 29009 bp in length.
Location/Qualifiers
1. 215734

FEATURES
Source

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-9J18"
/clone_lib="RPCI mouse BAC library 23"
BASE COUNT      54325 a 53055 c 52767 g 53583 t 2004 others
ORIGIN

```

Query Match	64.8%;	Score 300.6;	DB 2;	Length 215734;
Best Local Similarity	83.7%;	Pred. No. 1.3e-72;		
Matches 389;	Conservative	0;	Mismatches 69;	Indels 7;
				Gaps 4;

OY	2	aagaagaaagagaatggtgacacgatgttgactgcctgatgaagtgtggcgttgtctcaaa	61
Db	8042	AAAGGAAGGGAATGGCACCAGGTACTGCTACCATTAAGGTTCCGCCTTACTTGCTCGGA	8101
OY	62	agttctgcgagattgaacygctctctgtaattttagccaagagacacgcttgygaaccagy	121
Db	8102	AATTACGTGAGATTCAGCGGCTCTCTCGA-TTGAGCCTAGGGCACGCCCTGAGAAGCTGCGG	8160
OY	122	tgaacctacaagaagcccggaatctccgcgagaaattcaagltgttcttcctctccacct	181
Db	8161	TGACCTCACAGAGCGCGGAGTCTCCGGCAGAAAGTTCAAGTGTGTTTTCTCTTTCCACCCT	8220
OY	182	tcttcaggaacttcogaaatccgcctctccggtgaagtcagcatagcgtctgcgtcac	241
Db	8221	TTCCAGGAGACTCCGAGAACCCCGCCTCTCCGGTGACGTACACACAGCGCGCGCTCAGGG	8280
OY	242	tataaactcccgggtgactgltgttggcgagattgaactcaagltcgcagctgttgaagat	301
Db	8281	TATAAAGCCGCGTGACGATGTTGGCGCAGATTGAGTCAAGCTCTGAGTTGTGGAAGAT	8340
OY	302	tacatgcgagaccgccggcgactccgcataccctt----gccggagacgcttgcgaca	357
Db	8341	TACATGCGATATCCCGCGCGAACCCCGCATCCCTTGCCGGCCGGACAGCCCTTGCGTACA	8400
OY	358	gccccgtgagacatcaagtcctcccgaggcccccaagccttgaggcgacatgaacygcttgcct	417
Db	8401	GCCGTGAAACAT-GCGTCCCCGAGAGCCCACGCGCTGAGGGCGACATGAACCCGCTGCGCTT	8459
OY	418	tgaagagcaatccggaagcccaagatcgcttcttggcacaaccgaaccgy	462
Db	8460	C CGAGCAGTCCGAGCCACGATCGC-TTTGGCAACCGAACC GG	8503

```

RESULT      3
LOCUS       MMU83984
DEFINITION  MMU83984      216 bp      DNA      ROD      07-JUL-1998
            Mus musculus apoptosis associated protein (GADD34) gene, promoter
            sequence.
ACCESSION   U83984
VERSION     U83984.1  GI:3258621
KEYWORDS
SOURCE      house mouse.
            Mus musculus
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 216)
            Hollander,M.C., Zhan,Q., Bae,I. and Fornace,A.J. Jr.
            Mammalian GADD34, an apoptosis- and DNA damage-inducible gene
            J. Biol. Chem. 272 (21), 13731-13737 (1997)
JOURNAL     97298078
MEDLINE     2 (bases 1 to 216)
REFERENCE   Hollander,M.C. and Fornace,A.J. Jr.
            Direct Submission
            Submitted (07-JAN-1997) LMP, NIH, Bldg. 37, Room 5C09, Bethesda, MD
            20892, USA
FEATURES
            Location/Qualifiers
                1..216
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                <1..>216
                /gene="GADD34"

```

	/note="growth arrest and DNA damage-inducible gene 34; Myd116 gene"
promoter	<1. .>216
	/gene="GADD34"
GC_signal	135. .146
	/gene="GADD34"
misc_signal	149. .158
	/gene="GADD34"
	/note="ATF/CRE site"
TATA_signal	177. .182
	/gene="GADD34"
BASE COUNT	40 a 66 c 63 g 47 t
ORIGIN	

Query Match	33.28;	Score 154;	DB 10;	Length 216;
Best Local Similarity	85.58;	Pred. No. 7.1e-32;		
Matches 183; Conservative	0;	Mismatches 30;	Indels 1;	Gaps 1;

QY	68	gcgagattga	cgctctct	tgattgag	ccaaggac	acgcctgg	gaaagcca	cgtygacct	127		
Db	4	GTGAGATCG	AGCGCTCT	CTCGA-TTG	AGCCTA	GGGCAC	GCCTGG	GAAGCTG	CGGTGACCT	62	
QY	128	cacaagccc	ggaatctc	gcgagaa	atctcag	tgtytgtt	ctctctct	ctccacctt	ctca	187	
Db	63	CACAGAGCC	CGAGTCC	CGGAGAA	GTTCAG	TGTGTTT	TCTCTCT	TTCCAC	CTTCCCA	122	
QY	188	gggacttc	cgaaactc	cgccctc	tcgcg	tgacgtc	agcatag	cgcctgc	gtcagact	ataaa	247
Db	123	GGGACTTCC	GAGACCC	CGCCTCT	CCGGT	GAGCTCA	GCACAC	AGCCCC	CGCTC	AGGGTATAAA	182
QY	248	ctcccggt	gatacgt	gttgcg	cagattg	actca					281
Db	183	AGCCGCGT	GGACGAT	GTTGG	CGCAGAT	TGAGTCA					216

RESULT	4
LOCUS	MMMDPRMR
DEFINITION	MMMDPRMR 2275 bp mRNA ROD 12-SEP-1993
ACCESSION	Mouse myeloid differentiation primary response mRNA encoding Myd116 protein.
KEYWORDS	X51829
SOURCE	Myd116 gene; myeloid differentiation primary response gene. house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 2275)
TITLE	Liebermann,D.A.
JOURNAL	Direct Submission
REFERENCE	Submitted (12-FEB-1990) Liebermann D.A., University of Pennsylvania School of Medicine, Dept of Biochemistry & Biophysics, Philadelphia PA 19104-6059, USA
AUTHORS	2 (bases 1 to 2275)
TITLE	Lord,K.A., Hoffman-Liebermann,B. and Liebermann,D.A.
JOURNAL	Sequence of Myd116 cDNA: a novel myeloid differentiation primary response gene induced by IL6
FEATURES	Nucleic Acids Res. 18 (9), 2823 (1990)
SOURCE	90251472
	Location/Qualifiers
	1..2275
	/organism="Mus musculus"
	/strain="SL"
	/db_xref="taxon:10090"
	/clone_lib="lambda ZAP"
	/clone="116"
	229..2202
	/note="Myd116 protein (AA 1-657)"
	/codon_start=1
	/protein_id="CAA36128.1"
	/db_xref="GI:53041"
	/db_xref="SWISS-PROT:P17564"


```
/translation="MAPSPRPQHVHLWRDAHNFFYLLSPLMGLLSRAWSRLRGPEVPEA
WLAKTYTGADQIEAAALITPTPVSGNLLPHGETEESGSPQSOAQRCLVEAESPP
ETWGLSNVDEYNNAKPGODDLKRKEMERTAGKATLQAGADKRLGEVVARREEGVAE
PAYPTSOLEGGPAENEGDEETVKTYYQASASIAIPGYKPTPVPLGEAEHQATEKGT
ENKADPSNSPSSGSHSRAMEYYSRKPKQEGEAKVEAHAGQGHPCRNAAEAEGGPET
TFVCTGNAFLKAWVYRPGEDTEEDNSDSAEEDTAQTGATPHTSAFLKAWYRPGE
DTEEDSDSDSAEDTAQTGATPHTSAFLKAWYRPGEDTEEDNSDSAEEDTAQTG
ATPHTSAFLKAWVYRPGEDTEEDNSDSAEEDTAQTGATPHTSAFLKAWYRPGEDT
EDDTEEDSENVAIPGDSSETADSSQSPCLQPORCLPGEKTRGSEEPPLFQVAFYLP
EKPSBPWAAPKPLRLQRLRLFKAPTRDQDPEIPLKARKVHFAEKVTVHFLAVWAGP
AQAARRGPWEQFARDRSRFAARIAQAEKILGPYLPDPSRARAWARLRNPISLPQSEPRS
SSEATPLTQDVTTPSPPLPSETPPSPSLYLGRRG"
2275
```

polya_site

/note="polya site"

BASE COUNT 577 a 652 c 636 g 410 t

ORIGIN

Query Match

Best Local Similarity 29.4%; Score 136.6; DB 10; Length 2275;

Matches 162; Conservative 0; Mismatches 19; Indels 4; Gaps 1;

QY 282 gtgcgcagctgtgtgaagattacatgcagagaccccgcgactcgcacccctt---g 337

Db 1 GCTCTGAGTTGTGTGAAGATTACATGCGATATCCCGCGACCCCGCATCCCTTGCCGG 60

QY 338 ccgggacagcctttgcgacagcccggtgagacatcacgtcccgagcccccagcctgaggg 397

Db 61 CCGGGACAGCCTTTGCTTACAGCCTGTGAACAATTGGCTCCCGAGCCCGACGCTGAGGG 120

QY 398 cgacatgaacgcgctgtgccttgagagcaatccgagccacgacgtcctttggcaacgga 457

Db 121 CGACATGAACCCCGCTGGCTTCGCGAGCAGCTCCGACCCACGATCGCTTTGGCAACGAGA 180

QY 458 accgg 462

Db 181 ACCGG 185

RESULT 5
AC024740/c
LOCUS AC024740 179538 bp DNA HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 19 clone RP11-801D6, WORKING DRAFT
SEQUENCE, 17 unordered pieces.
AC024740
AC024740.3 GI:8570405
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 179538)
JOURNAL Waterston,R.H.
REFERENCE The sequence of Homo sapiens clone
AUTHORS Unpublished
TITLE 2 (bases 1 to 179538)
JOURNAL Waterston,R.H.
REFERENCE Direct Submission
AUTHORS Submitted (01-MAR-2000) Genome Sequencing Center, Washington
TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL MO 63108, USA
COMMENT On Jun 17, 2000 this sequence version replaced gi:7263917.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0801D06
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 97% of reads

Chemistry: Dye-terminator Big Dye; 3% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 173830 bases at least Q40
Consensus quality: 175557 bases at least Q30
Consensus quality: 176446 bases at least Q20
Insert size: 242000; agarose-fp
Insert size: 177938; sum-of-contigs
Quality coverage: 5.32 in Q20 bases; agarose-fp
Quality coverage: 6.01 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1801: contig of 1801 bp in length
* 1802 1901: gap of unknown length
* 1902 4641: contig of 2740 bp in length
* 4642 4741: gap of unknown length
* 4742 9679: contig of 4938 bp in length
* 9680 9779: gap of unknown length
* 9780 13392: contig of 3613 bp in length
* 13393 13492: gap of unknown length
* 13493 17164: contig of 3672 bp in length
* 17165 17264: gap of unknown length
* 17265 21510: contig of 4246 bp in length
* 21511 21610: gap of unknown length
* 21611 27089: contig of 5479 bp in length
* 27090 27189: gap of unknown length
* 27190 35636: gap of 8447 bp in length
* 35637 35736: gap of unknown length
* 35737 45537: contig of 9801 bp in length
* 45538 45637: gap of unknown length
* 45638 56451: contig of 10814 bp in length
* 56452 56551: gap of unknown length
* 56552 66802: contig of 10251 bp in length
* 66803 66902: gap of unknown length
* 66903 79118: contig of 12216 bp in length
* 79119 79218: gap of unknown length
* 79219 94355: contig of 15137 bp in length
* 94356 94455: gap of unknown length
* 94456 111274: contig of 16819 bp in length
* 111275 111374: gap of unknown length
* 111375 130600: contig of 19226 bp in length
* 130601 130700: gap of unknown length
* 130701 150278: contig of 19578 bp in length
* 150279 150378: gap of unknown length
* 150379 179538: contig of 29160 bp in length.

FEATURES
Location/Qualifiers

source

1. 179538
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-801D6"
1. 1801
/note="assembly_name:Contig13"
1902. 4641
/note="assembly_name:Contig14"
4742. 9679
/note="assembly_name:Contig15"
9780. 13392
/note="assembly_name:Contig16"
13493. 17164
/note="assembly_name:Contig17"
17265. 21510
/note="assembly_name:Contig18"
21611. 27089
/note="assembly_name:Contig19"
27190. 35636
/note="assembly_name:Contig20"

```
misc_feature      35737..45537
                  /note="assembly_name:Contig21"
misc_feature      45638..56451
                  /note="assembly_name:Contig22"
misc_feature      56552..66802
                  /note="assembly_name:Contig23"
misc_feature      66903..79118
                  /note="assembly_name:Contig24"
misc_feature      79219..94355
                  /note="assembly_name:Contig25"
misc_feature      94456..111274
                  /note="assembly_name:Contig26"
misc_feature      111375..130600
                  /note="assembly_name:Contig27"
misc_feature      130701..150278
                  /note="assembly_name:Contig28"
                  clone_end:SP6
                  vector_side:left"
misc_feature      150379..179538
                  /note="assembly_name:Contig29"
BASE COUNT      42230 a 47023 c 46289 g 42392 t 1604 others
ORIGIN
```

```
Query Match      25.1%; Score 116.4; DB 2; Length 179538;
Best Local Similarity 65.6%; Pred. No. 9e-22;
Matches 231; Conservative 0; Mismatches 116; Indels 5; Gaps 4;
```

```
QY 116 ccacggtgacctcacaagcccggaattccgcgagaatttcagtggtttcctct 175
    ||| ||||| | ||| | ||||| ||||| ||| ||| ||| |||
Db 156227 CCGCGGTGACGTTGCCAAAGCTGGAATCTCCGAGAGTCCCTGTTACTTCCACTTCC 156168

QY 176 ccaaccttctcagggacttccgaactccgcctctccg-gtgcagtcagcatagcgct-g 233
    | |||| | ||| | ||||| ||||| ||| ||||| ||| |||
Db 156167 CACCCTTCGGGTTGCGGTTCTCGAACCCTCTCTTCGTGACGTCAGCACGCCGGCG 156108

QY 234 cgtcagactataaactcccggtgatcggtgttgccgagattgactcaagttcgca-gctt 292
    || || ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156107 CGGTAGGCTATTAAGACCTAGTGCCATGTGTTCGCTGCTCTTATCGGTTCCCATCCCA 156048

QY 293 gtggaagattacatcgagagaccccgcgagctccgcattcccttgcgggagacgcctttg 352
    || | ||| ||||| ||||| ||| ||| ||||| ||||| ||| |||
Db 156047 GTTGTGATCTTATGCAAGACGCTGCACGACCCCGCGCTTGTGCCACGGCAGCTTG 155988

QY 353 cgacagcccggtgagacatcacgt--ccccgagcccaagcctgagcgagacatgaacgcg 410
    | |||| | ||| ||| ||| ||| ||||| ||||| ||| ||||| |||
Db 155987 AGGCAGCCGGAGATACTCTGAGTTACTCGAGCCCGACGCTGAGGGTGAGATGAACGCG 155928

QY 411 ctggccttgagagcaatccgcgacccacgactgcttttgcaacgcgaacgcgg 462
    ||||| | | ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 155927 CTGGCCTCCCTAACCGTCCGAGACCTGTGATCGCTTCTGGCAGACCGAACCGG 155876
```

RESULT 6
AC008749_3
WPCOMMENT

```
Sequence split into 5 fragments LOCUS AC008749 Accession AC008749
Fragment Name      Begin      End
AC008749_0         1      110000
AC008749_1        100001    210000
AC008749_2        200001    310000
AC008749_3        300001    410000
AC008749_4        400001    467420
Continuation (4 of 5) of AC008749 from base 300001 (AC008749 Homo sapiens chromosome 19
```

```
Query Match      24.7%; Score 114.8; DB 2; Length 110000;
Best Local Similarity 65.3%; Pred. No. 2.6e-21;
Matches 230; Conservative 0; Mismatches 117; Indels 5; Gaps 4;

QY 116 ccacggtgacctcacaagcccggaattccgcgagaatttcagtggtttcctctct 175
    ||| ||||| | ||| | ||||| ||||| ||| ||| ||| |||
Db 49833 CCGCGGTGACGTTGCCAAAGCTGGAATCTCCGCGAGAGTCCGTGTTACTTCCACTTCC 49892
```

```
QY 176 ccaccttctcagggacttccgaactccgcctctccg-gtgcagtcagcatagcgct-g 233
    | |||| | | ||||| ||||| ||||| ||||| ||| |||
Db 49893 CACCCTTCGGGTTGCGGTTCTCGAACCCTCTCTTCTGTCGACGTCAGCACGCCGGGCG 49952

QY 234 cgtcagactataaactcccggtgatcggtgttgccgagattgactcaagttcgca-gctt 292
    || || ||||| || | ||| || | | ||| || |
Db 49953 CGGTAGGCTATTAAGCCTTAGTGCCATGTGTTCGCTCTTATCGGTTCCCATCCCA 50012

QY 293 gtggaagattacatgagagaccccgcgagctccgcattcccttgcgggagacgcctttg 352
    || | ||| ||||| ||||| ||||| ||| ||| ||| ||| |||
Db 50013 GTTGTGATCTTATGCAAGACGCTGCACGACCCCGCGCTTGTGCCACGGCAGCTTG 50072

QY 353 cgacagcccggtgagacatcacgt--ccccgagcccaagcctgagggcgacatgaacgcg 410
    | |||| | ||| | || | ||||| ||||| ||| ||||| |||
Db 50073 AGGCAGCCGGAGATACTCTGAGTTACTCGGAGCCCGACGCTGAGGGTGAGATGAACGCG 50132

QY 411 ctggccttgagagcaatccgcgacccacgactgcttttgcaacgcgaacgcgg 462
    ||||| | | ||||| ||||| ||||| ||||| ||||| |||||
Db 50133 CTGGCCTCCCTAACCGTCCGAGACCTGTGATCGCTTCTGGCAGACCGAACCGG 50184
```

RESULT 7

```
AC026803
LOCUS      AC026803 231450 bp DNA HTG 26-JUL-2001
DEFINITION Homo sapiens chromosome 19 clone CTD-2639E6, *** SEQUENCING IN
PROGRESS *** 2 ordered pieces.
ACCESSION  AC026803
VERSION    AC026803.5 GI:15022021
KEYWORDS   HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.
SOURCE     human.
ORGANISM   Homo sapiens
```

```
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       1 (bases 1 to 231450)
JOURNAL     DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE   Unpublished
AUTHORS     2 (bases 1 to 231450)
TITLE       DOE Joint Genome Institute.
JOURNAL     Direct Submission
COMMENT     Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            On Jul 26, 2001 this sequence version replaced gi:9256691.
```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1 67040: contig of 67040 bp in length

* 67041 67140: gap of unknown length

* 67141 231450: contig of 164310 bp in length.

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 843459
Center clone name: CTD-2639E6

Summary Statistics
Consensus quality: 228571 bases at least Q40
Consensus quality: 230251 bases at least Q30
Consensus quality: 230412 bases at least Q20
Estimated insert size: 226000; agarose-fp estimation
Estimated insert size: 230499; sum-of-contigs
estimation
Quality coverage: 7.98 in Q20 bases; agarose-fp

```

      estimation      Quality coverage: 7.83 in Q20 bases; sum-of-contigs
      estimation.
FEATURES
      source
      location/Qualifiers
      1..231450
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="19"
      /clone="CTD-2639E6"
BASE COUNT      55608 a 59790 c 60166 g 55786 t 100 others
ORIGIN

Query Match      24.7%; Score 114.8; DB 2; Length 231450;
Best Local Similarity 65.3%; Pred. No. 2.4e-21;
Matches 230; Conservative 0; Mismatches 117; Indels 5; Gaps 4;

QY      116 ccacggtgacctcacaagggccggaatctccgagaaatttcagtggtttctctct 175
      b 116775 CCGCCGTGACGTTGCAAAAGCTGAAATCTCCGCGAGAAGTCCGTCTTACTTCCACTTCC 116834

QY      176 ccaccttctcagggacttcgaaactccgcctctccg-gtgaagtcagcatagcgt-g 233
      Db 116835 CACCCTTCGGGTTGCGGTCTCGAACCCTCCGCTCTCTTCGTGACGTCAACACGCCGGCCG 116894

QY      234 cgtcagactataaactcccggtgatcgtgttggcgcagattgactcagtcgca-gctt 292
      Db 116895 CCGTAGGCTATAAAAGCCTAGTGCCCATGTGTTCGTGCTTATCGGTTCCCATCCCA 116954

QY      293 gtggaagattacatgcgagaccccgcgagctccgcattcccttgcgggacagcctt 352
      Db 116955 GTGTGTGATCTTATGCAAGACGCTGCACGACCCCGCCGCTTGTGCGCACGCACTTG 117014

QY      353 cgcacagcccggtgagacatcaagt--ccccgagcccccaagcctgagggcgacatgaacgcg 410
      Db 117015 AGGCAAGCCGGAGATACTCTGAGTTACTCGGAGCCCGACGCTGAGGGGTGAGATGAACGCG 117074

QY      411 ctggccttgagagcaatccggaacccaagatcgcttcttgcaaacccaacgcg 462
      Db 117075 CTGGCCTCCCTAACCGTCCGGACCTGTGATCGCTTGTGGCAGACCGAACCCGG 117126

RESULT      8
CRUGADD34X      2088 bp      mRNA      ROD      29-JUN-1994
LOCUS      CRUGADD34X      2088 bp      mRNA      complete cds.
DEFINITION      Hamster Gadd34 mRNA, complete cds.
ACCESSION      L28147
VERSION      L28147.1 GI:452489
KEYWORDS      acidic protein.
SOURCE      Cricetulus longicaudatus cDNA to mRNA.
ORGANISM      Cricetulus longicaudatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE      1 (bases 1 to 2088)
AUTHORS      Zhan,Q., Lord,K.A., Alamo,I.Jr., Hollander,M.C., Carrier,F.,
      Ron,D., Kohn,K.W., Hoffman,B., Liebermann,D.A. and Fornace,A.J.Jr.
TITLE      The gadd and MyD genes define a novel set of mammalian genes
      encoding acidic proteins that synergistically suppress cell growth
JOURNAL      Mol. Cell. Biol. 14, 2361-2371 (1994)
MEDLINE      94187707
FEATURES
      source
      Location/Qualifiers
      1..2088
      /organism="Cricetulus longicaudatus"
      /db_xref="taxon:10030"
      /cell_line="Chinese hamster ovary-K1"
      215..1987
      /gene="Gadd34"
      215..1987
      /gene="Gadd34"
      215..1987
      /note="Gadd34 is the hamster homolog of the mouse MyD116
      cDNA; contains multiple PEST sequences implicated in rapid
      protein turnover"
```

```

/codon_start=1
/protein_id="AAA36983.1"
/db_xref="GI:452490"
/translation="MAPSPROHILLMRDAHSFHLSPLMGFLSRWSRLRVPAPEP
WPAETVTGADQIEADAHAPAPLPVPENNPPQGEAEESGTPBEKGAAQGPCLDVQANSSP
PETLGLSDDDKOGODPREQGRAHTAGLPILSPGIOSADKSLGEVVAEGEVTGLAY
PTSHWEGCPSSEEDGEDTGVKAFRASADSPGHKSSVYCPEAEHQATEKEQTEKKA
DEPSSPSGSHRAWYCSKQGEADPEPHRAGKYQLCONAEAESEEAQVSSLSVSSG
NAFLKAWVYRPGEDTEDDDSDWGSAAEEGKALSSPTSPHEFLKAWVYRPGEDTEDD
DDSDWGSAAEEGKALSSPTSPHEFLKAWVYRPGEDTEDDQSDWGSAAEKDGLAOTFA
TPHTSAFLKTWCCPGEDTEDDCEVVVPEDSEADPDKSPSHEAQCLPGEQTEGLY
EAEHSLFOVAFYLPGEKPAPEWTA PKPLRLQRRLLRLTPQDQDPETPLRAKVHF
SENVTVHFLAVWAGPAQARGPWEOLARDRSFRARIAQAEKIGPYLTTPAFARAW
ARLGNPSLPLALEPICDHTRFFPSQ"
```

```

Query Match      24.6%; Score 114.2; DB 10; Length 2088;
Best Local Similarity 83.0%; Pred. No. 5.5e-21;
Matches 142; Conservative 0; Mismatches 28; Indels 1; Gaps 1;

QY      293 gtggaagattacatgcgagaccccgcgcg-actccgcatcccttgcgggacagcctt 351
      Db 1 GTGGAAGATTGCATGCGACACCCCGCGCCGCCCGCATCCCTTGTGCGGACGCTCCCT 60

QY      352 gcgcacagcccggtgagacatcaagtcgcccgagcccccaagcctgagggcgacatgaacgcg 411
      Db 61 GCGGCCGTCTGTGAGACCTCACGTCCTCCCGAGCCCGCAGCCTGAGGGCGACATGAACGCTC 120

QY      412 tggccttgagagcaatccggaacccaagatcgcttcttgcaaacccaacgcg 462
      Db 121 TGCCCTTGCGAGACCTCCGGACCGAGACCGCTTGTGGCCACAGAACCCGG 171
```

```

RESULT      9
AC064877/c      165166 bp      DNA      HTG      29-JUN-2001
LOCUS      AC064877/c      165166 bp      DNA      WORKING DRAFT SEQUENCE,
DEFINITION      Homo sapiens chromosome 2 clone RP11-782E3, WORKING DRAFT SEQUENCE,
      2 unordered pieces.
ACCESSION      AC064877
VERSION      AC064877.5 GI:14575806
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 165166)
AUTHORS      Waterston,R.H.
TITLE      The sequence of Homo sapiens clone
JOURNAL      unpublished
REFERENCE      2 (bases 1 to 165166)
AUTHORS      Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (23-APR-2000) Genome Sequencing Center, Washington
      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
      MO 63108, USA
COMMENT      On Jun 29, 2001 this sequence version replaced gi:14280310.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0782E03
```

```
----- Summary Statistics -----
Sequencing vector: M13; 49%
Sequencing vector: plasmid; 50%
Chemistry: Dye-primer ET; 47% of reads
Chemistry: Dye-terminator Big Dye; 53% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 164535 bases at least Q40
Consensus quality: 164782 bases at least Q30
Consensus quality: 164917 bases at least Q20
Insert size: 162000; agarose-1p
Insert size: 165066; sum-of-contigs
Quality coverage: 14.14 in Q20 bases; agarose-1p
Quality coverage: 13.88 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 163910: contig of 163910 bp in length
* 163911 164010: gap of unknown length
* 164011 165166: contig of 1156 bp in length.
Location/Qualifiers
source
  1. 165166
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="2"
    /clone="RP11-782E3"
    /note="assembly_name:Contig24"
    /misc_feature
      164011..165166
    /note="assembly_name:Contig4"
BASE COUNT 52669 a 30734 c 29916 g 51745 t 102 others
ORIGIN
Query Match 22.9%; Score 106.2; DB 2; Length 165166;
Best Local Similarity 61.2%; Pred. No. 5.9e-19;
Matches 237; Conservative 0; Mismatches 145; Indels 5; Gaps 4;
QY 72 gattgaagcgtctctgtgatttgagccaagacagcctgagaa-gccacggtgacctcac 130
Db 164397 GATTGACAGTTCGTTGTTGGAGGGGGTGGTGCACGCTCGGAACCTCCGCCGTGACGTTGC 164338
QY 131 aagggcccggaatctccgcgagaatttcagtgtgtttctctctccaccttctcaggg 190
Db 164337 AAAAGCTGAATCTCCGCGAGAGAGTCTCTGTTACTTCCACTTCNNCCCTTCGGGTTGGC 164278
QY 191 acttcgaaactccgcctctccggtgacgtcagcatagcgct-gcgtcagactataaact 249
Db 164277 GTCTGAAACCCCGCTCTCTTCGTGACGTACGACACCGCGGGCGGTAGGCTATATAAG 164218
QY 250 cccggtgatcggtgttgscgagattgactcagtcgca-gcttgtgaagattacatgc 308
Db 164217 CCTAGTGGCCATTGTTGCTGCTCTTATCGGTCCATCCCAAGTTGTTGATCTTATGC 164158
QY 309 gagaccccgcgactccgcaccttgccgagacagccttgcagacagcccgtagac 368
Db 164157 AAGACGTCGACGACCCCGCGGCTTGTCCGACGCACTTGAGGACGCGGAGATAC 164098
QY 369 atcacgt--ccccagccccacgacctgagggcagcatgacgcgctgacctgagacaa 426
Db 164097 TCTGAGTACTCGGAGCCCGACGCGCTGAGGGTGAGATGAACGCGCTGCCCTTAACCG 164038
QY 427 tccgaccacagatcgctttgycaaa 453
Db 164037 TCCGACCTGTGATCGCTTCTGGCAGA 164011
```

```
RESULT 10
CGU83983 224 bp DNA 07-JUL-1998
LOCUS Cricetulus griseus apoptosis associated protein (GADD34) gene,
DEFINITION promoter sequence.
ACCESSION U83983
VERSION U83983.1 GI:3258620
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 224)
AUTHORS Hollander,M.C., Zhan,Q., Bae,I. and Fornace,A.J. Jr.
TITLE Mammalian GADD34, an apoptosis- and DNA damage-inducible gene
JOURNAL J. Biol. Chem. 272 (21), 13731-13737 (1997)
MEDLINE 97298078
REFERENCE 2 (bases 1 to 224)
AUTHORS Hollander,M.C. and Fornace,A.J. Jr.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1997) LMP, NIH, Bldg. 37, Room 5C09, Bethesda, MD
20892, USA
FEATURES
source location/Qualifiers
  1..224
    /organism="Cricetulus griseus"
    /db_xref="taxon:10029"
    <1..>224
    /gene="GADD34"
    /note="growth arrest and DNA damage-inducible gene 34"
    <1..>224
    /gene="GADD34"
    /note="ATF/CRE site"
    174..179
    /gene="GADD34"
    TATA_signal
BASE COUNT 46 a 67 c 63 g 48 t
ORIGIN
Query Match 20.7%; Score 96.2; DB 10; Length 224;
Best Local Similarity 73.5%; Pred. No. 6.3e-16;
Matches 164; Conservative 0; Mismatches 53; Indels 6; Gaps 3;
QY 67 tgcgagattgaagcgtctctgtgatttgagccaagacagcctgagaa-aggcaggtgac 125
Db 3 TTCGAGATTGACAGCTCGCAAGGTTGGGCCGAGACACGCGCTGGAAGTGGCAATGAC 62
QY 126 ctcaagaagcccggaatctccgcgagaaatttcagtggtttctctctccaccttct 185
Db 63 GTCACAAAGCCCGGAATTTCCGCGAGAGTTCCGTTGTT--CCCTGTCCACACTTTC 120
QY 186 cagggaactccgaaactccgcctctccggtgacgtcagcatagcgctgctcagactata 245
Db 121 CCGAGCTTCGGAAGCCCGCGCTCTCCGTTGACGTACGA---GGTCTCGCCAGGCTATA 177
QY 246 aactccgggtgacgtgttgscgagattgactcagtcgca 288
Db 178 AAAGCAGTGTGGCGCTTCGCGGCGAGATGAGCCACGTCCTCA 220
RESULT 11
HSU83981 2331 bp mRNA 07-JUL-1998
LOCUS Homo sapiens apoptosis associated protein (GADD34) mRNA, complete
DEFINITION cds.
ACCESSION U83981
VERSION U83981.1 GI:3258617
KEYWORDS human.
SOURCE
```


ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 2331)
JOURNAL	Hollander, M.C., Zhan, Q., Bae, I. and Fornace, A.J. Jr.
MEDLINE	Mammalian GADD34, an apoptosis- and DNA damage-inducible gene
REFERENCE	J. Biol. Chem. 272 (21), 13731-13737 (1997)
AUTHORS	97298078
TITLE	2 (bases 1 to 2331)
JOURNAL	Hollander, M.C. and Fornace, A.J. Jr.
REFERENCE	Direct Submission
AUTHORS	Submitted (07-JAN-1997) LMP, NIH, Bldg. 37, Room 5C09, Bethesda, MD
TITLE	20892, USA
FEATURES	
source	Location/Qualifiers
	1..2331
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="19"
	/map="19q13.2"
	1..2331
	/gene="GADD34"
	/note="growth arrest and DNA damage-inducible gene 34"
	223..2247
	/gene="GADD34"
	/function="growth arrest"
	/note="Gadd34"
	/codon_start=1
	/product="apoptosis associated protein"
	/protein_id="AAC25631.1"
	/db_xref="GI:3258618"
	/translation="MAPQAAPHQATPMRDAHPFLLSPVMGLSRMSRLRLGLPLP
	WLVEAVKGAALVEAGLEGEARTPLATPTWGRPEEEAEDSGPGEDETLGLTSS
	SLPEAMGLDDDDGMYGREATSVPRGGSQFADGQAPLSPSLIRTLQSDKNPGE
	EKAEEGVAAEEGVNKFSTYPPSHRECCPAVEEDEEA VKKEAHRISTALSPGSKPS
	TWVSCPGEEENQATEDKRTERSKGARTSVSPRSSGSDPRSWERYSGSEKEKAH
	KETGKGFAAPGPQSSAPQAPOLKSWCQPSDEEEGEVAKLGAEEKGEAECPCIP
	PSAFLKAWYVWPGEDTTEEEDEDEDEDESDSGSDEEEGEAAASSSTPATGVFLKSWYQ
	PGEDTEEEDEDESDTSAEDEREAEATSTPPASAFILKAWYVRPGEDTEEEDEDESDVS
	EDKEDDSEALGEAESDPHPSPDQRAHFRGWGVRPGKETEEEAEDMGEAEPCEPR
	VAIVVPEGKPPPPWAPRPLPLRLQRLKRPETPTHDPETPLKARKYRSEKVTVHF
	LAVWAGPAAQAQRQGPWEQLARDRSFRARRITQAQELSPCLTPAARARAWARLRNPPL
	APIPALQTLPPSSSVSPSSPVQTTPLSQAVATPPSRSSAAAAALDLGRG"
	1231..1332
	repeat_region
	1372..1473
	repeat_region
	1501..1602
	repeat_region
	1651..1752
	misc_feature
	1879..2088
	/gene="GADD34"
	/note="similar to HSV-1 ICP34.5 protein, SwissProt
	Accession Number P36313 and ASFV DP71L gene deposited
	under Genbank Accession Number U18466"
BASE COUNT	543 a 666 c 712 g 410 t
ORIGIN	
Query Match	17.5%; Score 81.2; DB 9; Length 2331;
Best Local Similarity	70.9%; Pred. No. 6.9e-12;
Matches 122; Conservative	0; Mismatches 48; Indels 2; Gaps 1;
QY	293 gtygaagattacatgcgagagaccccgcgactccgcattcttgcgagacagcttgg 352
Db	5 GTTGTGATCTTATGCAAGACGCTGCACGACCCCGCGCCTTGTCGCCACGGCATTG 64
QY	353 cgacagcccgtagacatcaagct--cccgagccccacagcctgagggcgacatgaacgcy 410
Db	65 AGGCAAGCCGGAGATACTCTGAGTTACTCTCGGAGCCCGACGCTGAGGGTGAGATGAACGCG 124
QY	411 ctgagccttagagcaatccgagaccacagatcgcttcttggcaaacggaacgcy 462
Db	125 CTGGCCTCCCTAACCCGTCGGACCTGTGATCGCTTCTGCGCAGACCGAACCGG 176

ARI00139	ARI00139	2344 bp	DNA	PAT	14-FEB-2001
LOCUS	Sequence 2	from patent US 6080558.			
DEFINITION	ARI00139				
ACCESSION	ARI00139				
VERSION	ARI00139.1	GI:12810587			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 2344)				
TITLE	Bandman,O., Ial,P., Shah,P. and Corley,N.C.				
JOURNAL	Polynucleotide encoding human growth regulator protein				
FEATURES	Patent: US 6080558-A 2 27-JUN-2000;				
	Location/Qualifiers				
source	1..2344				
BASE COUNT	556 a	667 c	709 g	412 t	
ORIGIN	/organism="unknown"				

Query Match	17.5%	Score 81.2	DB 6	Length 2344
Best Local Similarity	70.9%	Pred. No. 6.9e-12		
Matches 122	Conservative 0	Mismatches 48	Indels 2	Gaps 1
QY	293	gtggaagatcatatcgagagacccgcgcgactccgcacatccctttgccggagacagccttg	352	
Db	8	GTGTGATCATTTATGCAAGACGCTGCACGACCCCGCGCCCTTGTGCGCACGACCACTTG	67	
QY	353	cgacagcccgctgagacatcacgt--ccccgagccccacgcctgaggcgacatgaacgcg	410	
Db	68	AGGCAGCCCGAGATACTCTGAGTTACTCGAGCCCGACGCTGAGGGTGAGATGAACCGC	127	
QY	411	ctgccttgaagcaatccgcgacccacgatcgctttgcaaacgcgaacgcg	462	
Db	128	CTGGCCTCCCTAACCCGTCGCGACCTGTGATCGCTTCTGGCAGACCGAACCGG	179	
RESULT	13			
AK001361	2349 bp	mrna	PRI	22-FEB-2000
LOCUS				
DEFINITION	Homo sapiens CDNA FLJ10499 fis, clone NT2RP2000346, weakly similar to MYELOID DIFFERENTIATION PRIMARY RESPONSE PROTEIN MYD116.			
ACCESSION	AK001361			
VERSION	AK001361.1	GI:7022572		
KEYWORDS	oligo capping; fis (full insert sequence).			
SOURCE	Homo sapiens teratocarcinoma cell_line:NT2 CDNA to mRNA, clone_lib:NT2RP2 clone:NT2RP2000346.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (sites)			
	Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wagatsuma,M., Hosokiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuo,Y., Ninomiya,K. and Iwayanagi,T.			
TITLE	NEDO human cDNA sequencing project			
JOURNAL	Unpublished (2000)			
REFERENCE	2 (bases 1 to 2349)			
AUTHORS	Isogai,T. and Otsuki,T.			
TITLE	Direct Submission			
JOURNAL	Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)			
COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science,			

University of Tokyo.
 Location/Qualifiers
 source
 1..2349
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="NT2"
 /cell_type="teratocarcinoma"
 /clone="NT2RP2000346"
 /clone_lib="NT2RP2"
 /note="cloning vector: pME18SFL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction."
 241..2265
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAA91649.1"
 /db_xref="GI:7022573"
 /translation="MAPGQAPHQATPWRDAHPFLLSPVMGLSRAMSLRGLGLPLEPWLVEAVKGALVLEAGLEGEARTPLAIPHPWGRRPGEEAEDSGPEDRETLGLKTSSSLPEAWGLLDDDDMGYGEREATSVPRGQSGFADGQRAPLSPILLRLTQSGSDKNPGEKAEEGVAEEEGVNFKEFSYPSHRECCPAVEEDEEAVKKEAHRSTALSPGSKPSTWVSCGPEEENQATEDKRTKRCARKTSVSPRSSGSDPRSWYRSGEASEEKEKAHEETGGEAAPGPOSSAPAPORPOLKSWCQPSDEESEVKPLGAEKDGAECCPICPPSAFLKAWVWYMPGEDTEEEDEDEDESDSGSDEEGEAEASSSTPATGVFLKSWYQPGEDTEEEDESDTGSADEREAEATSASTPPASAFLKAWVYRPGEDTEEEDEVDVSDKEDDESEALGEAESDPIHPSHDQSAHFRGWYRPGKETEEEAEDMGEAEPCEPFRVAIVYGEKPPPPWAPPRPLRLRQRLRKRPETPTHDPDETPLKARKVRFSEKVTVHFLAVWAGPAQAAARQGPMWEQLARDRSRFARRIAQAQELSPCLTPAARARAWARLRNPPLAPIPALQTLPSSSVSPSVQVTPPLSQAVATPSRSSAAAAALLDLSGRG"

BASE COUNT 544 a 674 c 713 g 418 t
 ORIGIN

Query Match 17.5%; Score 81.2; DB 9; Length 2349;
 Best Local Similarity 70.9%; Pred. No. 6.9e-12;
 Matches 122; Conservative 0; Mismatches 48; Indels 2; Gaps 1;

Oy 293 gtggaagattacatgacgagaccccgcgagctccgcatcccttgcggyacagccttg 352
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 23 GTGTGTGATCTTATGCAAGACGCTGCACGACACCCCGCCCGCTTGTCTGCCACGGCACTTG 82

Oy 353 cgacagcccgtaagacatcaagcgt--ccccgagcccccacagccttgagggcgacatgaagcg 410
 | ||||| | ||| ||| | | | ||||| ||||| ||||| ||| ||||| ||| ||||| |||
 Db 83 AGGCAGCCCGAGATACTCTGAGTTACTCGGAGCGCCGACGCTGAGGGGTGAGATGAACCGC 142

Oy 411 ctggccttgagagcaatccgagccacagatcgcttctgcaaacggaacgg 462
 ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 143 CTGGCCTCCCTAACCGTCCGACCTGTGATCGCTTCTGGCAGACCGAACCGG 194

RESULT 14
 BC003067 2358 bp mRNA PRI 12-JUL-2001
 LOCUS
 DEFINITION Homo sapiens, growth arrest and DNA-damage-inducible 34, clone
 MGC:1985 IMAGE:3507556, mRNA, complete cds.
 ACCESSION BC003067
 VERSION BC003067.1 GI:13111759
 KEYWORDS MGC.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2358)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (12-FEB-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK
 COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadan@systemsbiology.org
 Anup Madan, Rachel Dichtoff, Jessica Fahey, Stephanie Ford, Julia
 Greene, Mark Ketteman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://Image.lnl.gov>
 Series: IRAL Plate: 7 Row: a Column: 18
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 3258617.

FEATURES
 source
 1. 2358
 /organism="Homo sapiens"
 /db_xref="LocusID:23645"
 /db_xref="taxon:9606"
 /clone="MGC:1985 IMAGE:3507556"
 /tissue_type="Brain, neuroblastoma"
 /clone_id="NH_MGC_19"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 232. 2256
 /codon_start=1
 /product="growth arrest and DNA-damage-inducible 34"
 /protein_id="AAH03067.1"
 /db_xref="GI:13111760"
 CDS
 232. 2256
 /translation="MAPQAPOHATPWRDAHFFLLSPVGMGLSRAMRLRGLPLEP
 WLVEAVKGAALVEAGLEGEARTPLAIPHTWGRPRPEEAEDSGGEDRETGLKTS
 SLPEAWGLDDDGMYGEREATVPRGSGQFADGORAPLSPSLIRTLQSDKNPGE
 EKAEEGVAAEEGVNFESYPPSHRECCPAVEEEDDEAVKKEAHTSTALSPSGKPS
 TWVSCGEEENQATEDKRTERSKGARKTSVPRSSGSDPRSWEYRSGEASEKEKAH
 KETGKEAAPGPOSSAPARQPLKSMWCQPSDEEEGEVKALGAERDGEAECPPIPP
 PSAFKLAWVWPGEDEEEDDEEEDSDSGSDEEEGEAEASSTPATGVFLKSWYO
 PGEDEEEDSDTGAEDEREAEETSASTPPASAFKLAWVIRPGEDEEEDSDVDS
 EDKEDDSEALGEAESDPHSPHPDORAHFRMGYRPGKETEEEEAEDWGEAEPCEFR
 VAIVPGEKPPPPWAPRPLRLQRLKRPETPHDPETPLKARVRFSEKTVHF
 LAVWAGPAOAROGPEWQLARDRSFRARITQAEELSPLTPAARARWARLRNPPL
 APIALQTLPSSSVPSPPVQTTPLSQAVAVPSSSAAAAALDLSGRG"

BASE COUNT 563 a 668 c 717 g 410 t
 ORIGIN

Query Match 17.5%; Score 81.2; DB 9; Length 2358;
 Best Local Similarity 70.9%; Pred. No. 6.9e-12;
 Matches 122; Conservative 0; Mismatches 48; Indels 2; Gaps 1;

QY	293	gtggaagattacatgagagacccgcgcgactccgcatcccttgcggagacacgttg	352
Db	14	GTGTGTGATCTTATGCAAGACGCTGCACAGACCCCGCGCTGTGCGCACAGGCACATTG	73
QY	353	cgacagcccgtagacatcacgt--cccgagcccccacgcgctgagggcgacatgaacgg	410
Db	74	AGGCAGCCCGAGATACTCTGAGTTACTCGAGACCCGACGCGCTGAGGGTGAGATGAACGCG	133
QY	411	ctggccctgagagcaatccgagccacgacgccttctggcaaacggaacgg	462
Db	134	CTGGCCCTCCCTAACCGCTCCGGACCTGTGATCGCTTCTGGCAGACCGAACCGG	185

RESULT 15
 HSU83982
 LOCUS
 DEFINITION Homo sapiens apoptosis associated protein (GADD34) gene, promoter
 sequence.
 ACCESSION U83982
 VERSION U83982.1 GI:3258619
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 233)
AUTHORS Hollander,M.C., Zhan,Q., Bae,I. and Fornace,A.J. Jr.
TITLE Mammalian GADD34, an apoptosis- and DNA damage-inducible gene
JOURNAL J. Biol. Chem. 272 (21), 13731-13737 (1997)
MEDLINE 97298078
REFERENCE 2 (bases 1 to 233)
AUTHORS Hollander,M.C. and Fornace,A.J. Jr.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1997) LMP, NIH, Bldg. 37, Room 5C09, Bethesda, MD 20892, USA

FEATURES
source location/Qualifiers
1..233
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.2"
<1..>233
/gene="GADD34"
/note="growth arrest and DNA damage-inducible gene 34"
promoter <1..>233
/gene="GADD34"
GC_signal 142..153
/gene="GADD34"
misc_signal 157..165
/gene="GADD34"
/note="ATF/CRE site"
TATA_signal 186..191
/gene="GADD34"
BASE COUNT 39 a 69 c 63 g 62 t
ORIGIN

Query Match 9.8%; Score 45.4; DB 9; Length 233;
Best Local Similarity 61.1%; Pred. No. 0.064;
Matches 107; Conservative 0; Mismatches 66; Indels 2; Gaps 2;
QY 116 ccacggtgacctcacaaagcccggaatctccgcgagaatttcagtgtgtttcctct 175
Db 58 CCGCCGTGACGTGCAAAAGCTGGAATCTCCGGAGAGTCCCTGCTTACTTCCACTTCC 117
QY 176 ccaccttctcagggaactccgaaactccgcctctccg-gtgaagtcagcatagcgt-g 233
Db 118 CACCCTTCGGGTTGCGGCTCTGAAACCCCGCCTCTTTCGTGACGTCAAGCAGCGGGCG 177
QY 234 cgtcagaactataaactcccggtgatcgtgttgcgagatgactcagttcgca 288
Db 178 CGGTAGGCTATATAAAGCCTAGTGCCATTGTGTGCTCTTATCGGGTCCCA 232

Search completed: April 5, 2002, 09:18:12
Job time: 3315 sec